

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 13:31:22 ; Search time 11.81 Seconds  
(without alignments)  
30.703 Million cell updates/sec

Title: US-09-646-043-1

Sequence: 1 ASFTAGDTTAAQVPSIV 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents-AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCPUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	82	94.3	185	1 US-08-233-788A-61	Sequence 61, Appl
2	42	48.3	205	1 US-08-277-231A-5	Sequence 5, Appl
3	42	48.3	205	2 US-08-473-750-8	Sequence 8, Appl
4	42	48.3	205	2 US-08-477-326-8	Sequence 8, Appl
5	40	46.0	165	6 5217891-5	Patent No. 5217891
6	40	46.0	1732	2 US-08-570-311-10	Sequence 10, Appl
7	40	46.0	1732	2 US-08-353-485-10	Sequence 10, Appl
8	40	46.0	2628	2 US-08-570-311-14	Sequence 14, Appl
9	39	44.8	387	3 US-08-689-421-23	Sequence 23, Appl
10	39	44.8	387	4 US-08-689-421-23	Sequence 23, Appl
11	39	44.8	387	4 US-09-181-827A-23	Sequence 23, Appl
12	39	44.8	516	4 US-08-689-421-29	Sequence 29, Appl
13	39	44.8	516	4 US-09-181-827A-29	Sequence 29, Appl
14	39	44.8	516	4 US-08-417-174-92	Sequence 92, Appl
15	38	43.7	9	2 US-08-218-943-1	Sequence 1, Appl
16	38	43.7	647	1 US-08-218-943-1	Sequence 1, Appl
17	37	42.5	1876	2 US-08-619-554-2	Sequence 2, Appl
18	37	42.5	1895	2 US-08-619-554-4	Sequence 4, Appl
19	36	41.4	245	5 US-08-272-255-15	Sequence 15, Appl
20	36	41.4	245	5 PCP-US95-08565-15	Sequence 15, Appl
21	36	41.4	574	6 5378464-3	Patent No. 5378464
22	36	41.4	681	2 US-08-272-255-6	Sequence 6, Appl
23	36	41.4	681	2 US-08-272-255-9	Sequence 9, Appl
24	36	41.4	681	4 US-08-964-268-5	Sequence 5, Appl
25	36	41.4	681	4 PCP-US95-08565-6	Sequence 6, Appl
26	36	41.4	681	5 PCP-US95-08565-9	Sequence 9, Appl
27	36	41.4	808	3 US-08-804-439A-14	Sequence 14, Appl

28	36	41.4	808	3 US-08-720-229-14	Sequence 14, Appl
29	36	41.4	830	1 US-08-110-158-4	Sequence 4, Appl
30	36	41.4	830	5 PCT-US91-05059-2	Sequence 2, Appl
31	35	40.2	9	2 US-08-417-174-04	Sequence 9, Appl
32	35	40.2	105	2 US-08-659-251-9	Sequence 9, Appl
33	35	40.2	105	4 US-09-256-490-9	Sequence 9, Appl
34	35	40.2	105	5 PCT-US96-11445-9	Sequence 9, Appl
35	35	40.2	467	2 US-08-727-548-2	Sequence 2, Appl
36	35	40.2	855	2 US-08-619-554-6	Sequence 6, Appl
37	35	40.2	874	3 US-08-804-439A-15	Sequence 15, Appl
38	35	40.2	874	3 US-08-720-229-16	Sequence 16, Appl
39	34	39.1	507	2 US-08-409-122-2	Sequence 2, Appl
40	34	39.1	507	2 US-08-408-669-2	Sequence 2, Appl
41	34	39.1	520	4 US-09-000-016-7	Sequence 7, Appl
42	34	39.1	734	4 US-09-000-016-4	Sequence 2, Appl
43	34	39.1	823	4 US-09-000-016-2	Sequence 2, Appl
44	34	39.1	874	3 US-08-804-439A-15	Sequence 15, Appl
45	34	39.1	874	3 US-08-720-229-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1  
US-08-233-788A-61  
Sequence 61, Application US/08233788A  
Patent No. 5635617  
GENERAL INFORMATION:  
APPLICANT: Doran, James L.  
APPLICANT: Kay, William W.  
APPLICANT: Collinson, Karen S.  
APPLICANT: Clouthier, Sharon C.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION  
NUMBER OF SEQUENCES: 61  
CORRESPONDENCE ADDRESSES:  
ADDRESS: Seed and Berry  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: U.S.A.  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/233,788A  
FILING DATE: 26-APR-1991  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: King, Joshua  
REGISTRATION NUMBER: 35,570  
REFERENCE/DOCKET NUMBER: 920043.403C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
TELEX: 3723836 SEDANBERY  
SEQUENCE CHARACTERISTICS:  
LENGTH: 185 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-233-788A-61  
Query Match 94.3%; Score 82; DB 1; Length 185;  
Best Local Similarity 94.4%; Pred. No. 5.8e-07;  
Matches 17; Conservative 1; Mismatches 0; Gaps 0;  
OY 1 ASFTAGDTTAAQVPSIV 18

DB 65 ASFTAGTAAQVPSIV 82

## RESULT 2

US-08-277-231A-5  
Sequence 5, Application US/08277231A  
Patent No. 5643725

GENERAL INFORMATION:  
APPLICANT: Green, Bruce A.  
APPLICANT: Brinton, Charles C.  
TITLE OF INVENTION: Sequence and Analysis of LKP Pillin  
Patent No. 5643725

TITLE OF INVENTION: Structural Genes and The LKP Pill Operon of No. 5643725typab  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive

CITY: Lexington  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/277,231A

FILING DATE: 19-JUL-1994  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Carroll, Alice O.  
REGISTRATION NUMBER: 32,542

REFERENCE/DOCKET NUMBER: ACC94-02  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540

INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:

LENGTH: 205 amino acids  
TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-277-231A-5

Query Match 48.3%; Score 42; DB 1; Length 205;  
Best Local Similarity 52.9%; Pred. No. 5.3;

Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 ASFTAGTAAQVPSI 17

DB 96 ANLSHAGTAAQVPSI 112

## RESULT 3

US-08-473-750-8  
Sequence 8, Application US/08473750  
Patent No. 5834187

GENERAL INFORMATION:  
APPLICANT: Green, Bruce A.

APPLICANT: Brinton, Jr., Charles C.  
TITLE OF INVENTION: Sequence and Analysis of LKP Pillin

Patent No. 5834187  
Patent No. 5834187 5786143

TITLE OF INVENTION: Structural Gene and the LKP Pill Operon of No. 5834187 5786143

NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: US

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/473,750

FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/277,321

FILING DATE: 19-JUL-1994  
ATTORNEY/AGENT INFORMATION:

NAME: Carroll, Alice O.  
REGISTRATION NUMBER: 33,542

REFERENCE/DOCKET NUMBER: ACC94-02B  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540

INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:

LENGTH: 205 amino acids  
TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-473-750-8

Query Match 48.3%; Score 42; DB 2; Length 205;  
Best Local Similarity 52.9%; Pred. No. 5.3;

Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 ASFTAGTAAQVPSI 17

DB 96 ANLSHAGTAAQVPSI 112

## RESULT 4

US-08-477-326-8  
Sequence 8, Application US/08477326  
Patent No. 5968769

GENERAL INFORMATION:  
APPLICANT: Green, Bruce A.

APPLICANT: Brinton, Jr., Charles C.  
TITLE OF INVENTION: Sequence and Analysis of LKP Pillin

Patent No. 5968769  
TITLE OF INVENTION: Structural Gene and the LKP Pill Operon of No. 5968769typab

NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive

CITY: Lexington  
STATE: Massachusetts

COUNTRY: US  
ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,326

FILING DATE: 07-JUN-1995  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/277,231

FILING DATE: July 19, 1994

Thu Jun 21 08:38:31 2001

us-09-646-043-1.ra1

Page 3

ATTORNEY/AGENT INFORMATION:

NAME: Carroll, Alice O. 542  
REGISTRATION NUMBER: 33, 542  
REFERENCE/DOCKET NUMBER: ACC94-02A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 205 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-477-326-8

Query Match 48.3%; Score 42; DB 2; Length 205;  
Best Local Similarity 52.9%; Pred. No. 5.3;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 1 ASPTAGDTTAQVPSI 17  
DB 96 ANLSHAGQTAAPVPSI 112

RESULT 5

5217891-5  
Patent No. 5217891  
APPLICANT: BRAKE, ANTHONY J.; VAN DEN BERG, JOHAN A.  
TITLE OF INVENTION: DNA CONSTRUCTS CONTAINING A KLUYEROMYCES  
A FACTOR LEADER SEQUENCE FOR DIRECTING SECRETION OF HETEROLOGOUS  
POLYPEPTIDES  
NUMBER OF SEQUENCES: 23  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/507,398  
FILING DATE: 09-APR-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 78,551  
FILING DATE: 28-JUL-1987  
SEQ ID NO: 5  
LENGTH: 165  
5217891-5

Query Match 46.0%; Score 40; DB 6; Length 165;  
Best Local Similarity 53.3%; Pred. No. 9.1;  
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 4 TAIGDTTAQVPSI 18  
DB 24 TTTGDTTAQVPSI 38

RESULT 6

US-08-570-311-10  
Sequence 10, Application US/08570311  
GENERAL INFORMATION:  
APPLICANT: Proguiske-Fox, Ann  
APPLICANT: Tumwasorn, Somying  
APPLICANT: Lepline, Guyalaine  
APPLICANT: Han, Naiming  
APPLICANT: Lant, Joseph  
APPLICANT: Lant, Joseph  
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes  
TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ted W. Whitlock  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA

ZIP: 32606  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/570,311

FILING DATE:

CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/353,485

FILING DATE: 09-DEC-1994

CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/647,119

FILING DATE: 25-JAN-1991

CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/241,640

FILING DATE: 08-SEP-1988

ATTORNEY/AGENT INFORMATION:  
NAME: Whitlock, Ted W.  
REGISTRATION NUMBER: 36,965

REFERENCE/DOCKET NUMBER: DP15.C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (904) 375-8100

TELEFAX: (904) 372-5800

INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:

LENGTH: 1732 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-570-311-10

Query Match 46.0%; Score 40; DB 2; Length 1732;  
Best Local Similarity 37.5%; Pred. No. 1.3e+02;  
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 2 SPTAGDTTAQVPSI 17  
DB 1658 NITSADVTAQKPYTL 1673

RESULT 7

US-08-353-485-10  
Sequence 10, Application US/08353485  
GENERAL INFORMATION:  
APPLICANT: Proguiske-Fox, Ann  
APPLICANT: Tumwasorn, Somying  
APPLICANT: Lepline, Guyalaine  
APPLICANT: Han, Naiming  
APPLICANT: Lant, Joseph  
APPLICANT: Lant, Joseph  
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes  
TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ted W. Whitlock  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/353,485  
FILING DATE: 09-DEC-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/647,119  
FILING DATE: 25-JAN-1991  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/241,640  
FILING DATE: 08-SEP-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Whitlock, Ted W.  
REGISTRATION NUMBER: 36,965  
REFERENCE/DOCKET NUMBER: UP15.C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (904) 375-8100  
TELEFAX: (904) 375-8100  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1732 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-353-485-10

Query Match  
Best Local Similarity 46.0%; Score 40; DB 2; Length 1732;  
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 2 SFTAGDTAQQVPSI 17  
:|::|1111::  
Db 1658 NITSLADVTAKRYTL 1673

RESULT 8  
US-08-570-311-14  
Sequence 14, Application US/08570311  
Patent No. 5824791  
GENERAL INFORMATION:  
APPLICANT: Proguiske-Fox, Ann  
APPLICANT: Tumwasorn, Somying  
APPLICANT: Lepine, Guylaine  
APPLICANT: Han, Naiming  
APPLICANT: Lantz, Marilyn  
APPLICANT: Patel, Joseph  
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Ted W. Whitlock  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/570,311  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/353,485  
FILING DATE: 09-DEC-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/647,119  
FILING DATE: 25-JAN-1991  
CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/241,640  
FILING DATE: 08-SEP-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Whitlock, Ted W.  
REGISTRATION NUMBER: 36,965  
REFERENCE/DOCKET NUMBER: UP15.C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (904) 375-8100  
TELEFAX: (904) 375-8100  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2628 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-570-311-14

Query Match  
Best Local Similarity 46.0%; Score 40; DB 2; Length 2628;  
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 2 SFTAGDTAQQVPSI 17  
:|::|1111::  
Db 2554 NITSLADVTAKRYTL 2569

RESULT 9  
US-08-689-421-23  
Sequence 23, Application US/08689421  
Patent No. 6008029  
GENERAL INFORMATION:  
APPLICANT: Yaver, Debbie S.  
APPLICANT: Brown, Kimberly M.  
APPLICANT: Kauppinen, Sakari  
APPLICANT: Hakler, Torben P  
TITLE OF INVENTION: PURIFIED COPRINUS LACCASES AND NUCLEIC ACIDS  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: No. 6008029disk of No. 6008029th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/689,421  
FILING DATE: 9-AUG-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Harrington, James J.  
REGISTRATION NUMBER: 38,711  
REFERENCE/DOCKET NUMBER: 4554.204-WO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 387 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-689-421-23

Query Match

44.8%; Score 39; DB 3; Length 387;

Thu Jun 21 08:38:31 2001

US-09-646-043-1.rai

Best Local Similarity 58.3%; Pred. No. 36;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 AIGDTTQAVPFS 16  
1: 11:1111  
DB 222 AVDPSTSQLPFS 233

RESULT 10  
US-09-389-528-23  
Sequence 23, Application US/09389528

PATENT No. 6207430  
GENERAL INFORMATION:  
APPLICANT: Yaver, Debbie S.  
APPLICANT: Brown, Kimberley M.  
APPLICANT: Kauppinen, Sakari P  
APPLICANT: Halkier, Torben P  
TITLE OF INVENTION: PURIFIED COPRINUS LACCASES AND NUCLEIC ACIDS  
TITLE OF INVENTION: ENCODING SAME  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESSES:  
ADDRESS: No. 6207430, No. 6207430disk of No. 6207430th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/389,528  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/689,421  
FILING DATE: 9-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Harrington, James J.  
REGISTRATION NUMBER: 38,711  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 387 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-09-389-528-23

Query Match 44.8%; Score 39; DB 4; Length 387;  
Best Local Similarity 58.3%; Pred. No. 36;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 AIGDTTQAVPFS 16  
1: 11:1111  
DB 222 AVDPSTSQLPFS 233

RESULT 11  
US-09-181-827A-23  
Sequence 23, Application US/09181827A

PATENT No. 6242232  
GENERAL INFORMATION:  
APPLICANT: Yaver, Debbie S.  
APPLICANT: Brown, Kimberley M.  
APPLICANT: Kauppinen, Sakari P  
APPLICANT: Halkier, Torben P

TITLE OF INVENTION: Purified Coprinus Laccases And Nucleic  
Acids Encoding Same  
FILE REFERENCE: 4554.200-US  
CURRENT APPLICATION NUMBER: US/09/181,827A  
CURRENT FILING DATE: 1998-10-28  
PRIOR FILING DATE: 1995-08-25  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 23  
LENGTH: 387  
TYPE: PRT  
ORGANISM: Coprinus cinereus

US-09-181-827A-23

Query Match 44.8%; Score 39; DB 4; Length 387;  
Best Local Similarity 58.3%; Pred. No. 36;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 AIGDTTQAVPFS 16  
1: 11:1111  
DB 222 AVDPSTSQLPFS 233

RESULT 12  
US-08-689-421-29  
Sequence 29, Application US/08689421

PATENT No. 6008029  
GENERAL INFORMATION:  
APPLICANT: Yaver, Debbie S.  
APPLICANT: Brown, Kimberley M.  
APPLICANT: Kauppinen, Sakari P  
APPLICANT: Halkier, Torben P  
TITLE OF INVENTION: PURIFIED COPRINUS LACCASES AND NUCLEIC ACIDS  
TITLE OF INVENTION: ENCODING SAME  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESSES:  
ADDRESS: No. 6008029, No. 6008029disk of No. 6008029th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/689,421  
FILING DATE: 9-AUG-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Harrington, James J.  
REGISTRATION NUMBER: 38,711  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 516 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-689-421-29

Query Match 44.8%; Score 39; DB 3; Length 516;  
Best Local Similarity 58.3%; Pred. No. 50;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 AIGDTTAAQVPS 16  
1: 11:1111  
Db 306 AVDPPTSOLPS 317

## RESULT 13

US-09-389-528-29  
Sequence 29, Application US/09389528  
Patent No. 6207430  
GENERAL INFORMATION:  
APPLICANT: Yaver, Debbie S.  
APPLICANT: Brown, Kimberley M.  
APPLICANT: Kauppinen, Sakari  
APPLICANT: Halkier, Torben P.  
TITLE OF INVENTION: PURIFIED COPRINUS LACCASES AND NUCLEIC ACIDS  
TITLE OF INVENTION: ENCODING SAME  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSER: No. 62074300 No. 6207430disk of No. 6207430th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/389,528  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/689,421  
FILING DATE: 9-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Harrington, James J.  
REGISTRATION NUMBER: 38,711  
REFERENCE/DOCKET NUMBER: 4554.204-WO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 516 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-389-528-29

## Query Match

Best Local Similarity 44.8%; Score 39; DB 4; Length 516;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 AIGDTTAAQVPS 16  
1: 11:1111  
Db 306 AVDPPTSOLPS 317

## RESULT 14

US-09-181-827A-29  
Sequence 29, Application US/09181827A  
Patent No. 6242232  
GENERAL INFORMATION:  
APPLICANT: Yaver, Debbie S.  
APPLICANT: Brown, Kimberley M.  
APPLICANT: Kauppinen, Sakari  
APPLICANT: Halkier, Torben P.  
TITLE OF INVENTION: Purified Coprinus Laccases And Nucleic  
FILE REFERENCE: 4554.200-US

CURRENT APPLICATION NUMBER: US/09/181,827A  
CURRENT FILING DATE: 1998-10-28  
PRIOR APPLICATION NUMBER: 60/002,800  
PRIOR FILING DATE: 1995-08-25  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 29  
LENGTH: 516  
TYPE: PRT  
ORGANISM: Coprinus cinereus  
US-09-181-827A-29

Query Match  
Best Local Similarity 44.8%; Score 39; DB 4; Length 516;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 AIGDTTAAQVPS 16  
1: 11:1111  
Db 306 AVDPPTSOLPS 317

## RESULT 15

US-08-417-174-92  
Sequence 92, Application US/08417174  
Patent No. 5844075  
GENERAL INFORMATION:  
APPLICANT: KAWAKAMI, YUTAKA; ROSENBERG,  
TITLE OF INVENTION: MELANOMA ANTIGENS AND  
TITLE OF INVENTION: THEIR USE IN DIAGNOSTIC AND THERAPEUTIC  
NUMBER OF SEQUENCES: 126  
CORRESPONDENCE ADDRESS:  
ADDRESSER: MORGAN & FINNEGAN, L.L.P.  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/417,174  
FILING DATE: 05-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/231,565  
FILING DATE: 22-APR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: CAROL M. GRUPPI  
REGISTRATION NUMBER: 37,341  
REFERENCE/DOCKET NUMBER: 2026-4124US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 92:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9  
TYPE: amino acid  
STRANDEDNESS: Unknown  
TOPOLOGY: Unknown  
MOLECULE TYPE: Peptide  
US-08-417-174-92

Query Match  
Best Local Similarity 43.7%; Score 38; DB 2; Length 9;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Thu Jun 21 08:38:31 2001

us-09-646-043-1.rail

Page 7

OY 10 TAOVPSI 17  
| | | | |  
Db 2 TAOVPSV 9

Search completed: June 20, 2001, 13:31:40  
Job time: 18 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: June 20, 2001, 13:31:22 ; Search time 13.45 Seconds  
(without alignments)  
101.944 Million cell updates/sec

Title: US-09-646-043-1

Perfect score: 87  
Sequence: 1 ASFTAGDTTAAQVPSIV 18

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

219241

Total number of hits satisfying chosen parameters:  
Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: PIR.68:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	100.0	184	2 S20682	type 1 fimbrial pr
2	82	94.3	185	2 B28393	type 1 fimbrial pr
3	49	56.3	191	2 A64785	probable fimbrial
4	49	56.3	191	2 C85553	fimbrial protein h
5	47	54.0	171	2 B64785	ABC-transporter st
6	44.5	51.1	591	2 S65588	probable fimbrial
7	44	50.6	171	2 C85553	probable WD-repeat
8	44	50.6	931	2 B85062	fimbrial protein h
9	43	49.4	216	2 S54430	hypothetical prote
10	43	49.4	659	2 B96827	hypothetical prote
11	42	48.3	1248	2 B96827	probable tyrosine-
12	42	48.3	194	2 T41416	hypothetical prote
13	42	48.3	445	2 E70204	long polar fimbria
14	42	48.3	1146	2 A56271	conserved hypochet
15	41	47.1	177	2 A81725	fimbrial protein - Esc
16	41	47.1	215	2 S56540	fimbrial protein - Esc
17	41	47.1	215	2 S56540	fimbrial protein - Esc
18	41	47.1	215	2 S56540	fimbrial protein - Esc
19	41	47.1	504	2 F84219	DNA-binding protein
20	41	47.1	628	2 T47523	fimbrial morpholog
21	40	46.0	176	2 S56543	hypothetical prote
22	40	46.0	176	2 H86129	hypothetical prote
23	40	46.0	1419	2 T32970	lysine-specific cy
24	40	46.0	1732	2 T30836	hemagglutinin A -
25	40	46.0	2628	2 T28651	hypothetical prote
26	39	44.8	270	2 T43039	hypothetical prote
27	39	44.8	409	2 S77262	probable PPE prote
28	39	44.8	468	2 B70932	
29	39	44.8	468	2 B70932	

30	39	44.8	488	1 Q06B85	alkaline exonuclease
31	39	44.8	488	2 T44030	alkaline exonuclease
32	39	44.8	488	2 T44215	alkaline exonuclease
33	39	44.8	615	2 T39758	proline-serine ric
34	39	44.8	725	2 E86790	hypothetical prote
35	38	43.7	85	2 T17696	hypothetical prote
36	38	43.7	170	2 T39892	hypothetical prote
37	38	43.7	175	2 S06193	slag protein precu
38	38	43.7	175	2 S06193	slag protein precu
39	38	43.7	175	2 S06193	slag protein precu
40	38	43.7	175	2 S06193	slag protein precu
41	38	43.7	175	2 S06193	slag protein precu
42	38	43.7	175	2 S06193	slag protein precu
43	38	43.7	175	2 S06193	slag protein precu
44	38	43.7	175	2 S06193	slag protein precu
45	38	43.7	175	2 S06193	slag protein precu

## ALIGNMENTS

RESULT 1  
S20682 type 1 fimbrial protein fima - Salmonella typhi  
C:Species: Salmonella typhi  
C:Date: 20-Feb-1995 #sequence-revision 20-Feb-1995 #text-change 26-Aug-1999  
C:Accession: S20682  
R:Rossolini, G.M.; Muscas, P.; Chlesurini, A.; Sakta, G.  
submitted to the EMBL Data Library, March 1992  
A:Reference number: S20682  
A:Accession: S20682  
A:Molecule type: DNA  
A:Residues: 1-184 <ROS>  
A:Cross-references: EMBL:X65168; NID:G395382; PIDN:CAA46286.1; PID:G47667  
A:Experimental source: strain Stc4  
A:Gene: fima  
C:Superfamily: type 1 fimbrial protein  
C:Keywords: fimbria

Query Match 100.0% Score 87; DB 2; Length 184;  
Best Local Similarity 100.0% Pred. No. 8.6e-08;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASFTAGDTTAAQVPSIV 18  
DB 65 ASFTAGDTTAAQVPSIV 82

RESULT 2  
B28393 type 1 fimbrial protein precursor - Salmonella typhimurium  
C:Species: Salmonella typhimurium  
C:Date: 28-Aug-1989 #sequence-revision 28-Aug-1989 #text-change 15-Nov-1996  
C:Accession: B28393; A05121  
R:Purcell, B.K.; Pruckner, J.; Cleeg, S.  
J. Bacteriol. 169, 5631-5634, 1987  
A:Title: Nucleotide sequences of the genes encoding type 1 fimbrial subunits of Klebs  
A:Reference number: A91858; M01D:88058806  
A:Accession: B28393  
A:Molecule type: DNA  
A:Residues: 1-185 <PUN>  
A:Note: The authors translated the codon GCC for residue 18 as Gly, GTG for residue 3  
R:Maalen, K.; Sletten, K.; Froholm, L.O.; Valsanen, V.; Korhonen, T.K.  
FEMS Microbiol. Lett. 16, 149-151, 1983  
A:Reference number: A05121  
A:Accession: A05121  
A:Molecule type: protein  
A:Residues: 23-30, 'S', 32-45, 'X', 47-50 <MAA>  
C:Superfamily: type 1 fimbrial protein  
F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-185/Product: type 1 fimbrial protein #status predicted <MAT>



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 13:31:22 ; Search time 8.38 Seconds  
(without alignments)  
73.580 Million cell updates/sec

Title: US-09-646-043-1

Perfect score: 87

Sequence: 1 ASFTAIGDTTAQVFFSIV 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	100.0	184	1 FMA_SALTI	P37920 salmonella
2	82	94.3	185	1 FMT_SALTI	B55233 salmonella
3	82	94.3	185	1 FMA_SALTY	P37921 salmonella
4	49	56.3	191	1 SFMA_ECOLI	P77660 escherichia
5	47	54.0	171	1 SFMF_ECOLI	P38052 escherichia
6	43	49.4	216	1 HFD1_HAEIN	P45992 haemophilus
7	42	48.3	205	1 HFD2_HAEIN	P45993 haemophilus
8	41	47.1	178	1 LPFA_SALTY	P43660 salmonella
9	41	47.1	211	1 VPV_BPP2	P31340 bacteriophage
10	41	47.1	215	1 FIMI_ECOLI	P39284 escherichia
11	41	47.1	504	1 VLI_HPV31	P17388 human papill
12	40	46.0	176	1 FIME_ECOLI	P08189 escherichia
13	40	46.0	2628	1 HAGA_PORGI	P51845 porphyromon
14	39	44.8	488	1 EXON_HSV6U	P24447 human herpe
15	39	44.8	488	1 EXON_HSV62	P52448 human herpe
16	38	43.7	170	1 YFSD_BACSU	P42961 bacillus su
17	38	43.7	175	1 SFAG_ECOLI	P13429 escherichia
18	38	43.7	193	1 ISPZ_VITS1	Q9X50 vitreoscilli
19	38	43.7	1135	1 VGLM_HANTV	P08668 hantaan vir
20	38	43.7	1246	1 VP03_HSVSA	Q01000 herpesvirus
21	38	43.7	1501	1 SNO2_YEAST	P32568 saccharomyc
22	37	42.5	227	1 PGSA_MYCPN	P75520 mycoplasma
23	37	42.5	285	1 GLO4_YEAST	Q12320 saccharomyc
24	37	42.5	354	1 YAUQ_SCHPO	Q10169 schizosacch
25	37	42.5	367	1 FLGI_PSEPU	Q52082 pseudomonas
26	37	42.5	851	1 YDBG_SCHPO	Q10366 schizosacch
27	37	42.5	902	1 YC47_SCHPO	O14053 schizosacch
28	37	42.5	1499	1 CDR2_CANAL	P78595 candida alb
29	37	42.5	1876	1 GLS1_YEAST	P38631 saccharomyc
30	37	42.5	1895	1 GLS2_YEAST	P40989 saccharomyc
31	36	41.4	132	1 MPP6_PHLPR	P43215 phleum prat
32	36	41.4	174	1 CALB_NEUCR	P87072 neurospora
33	36	41.4	194	1 YRAH_ECOLI	P42913 escherichia

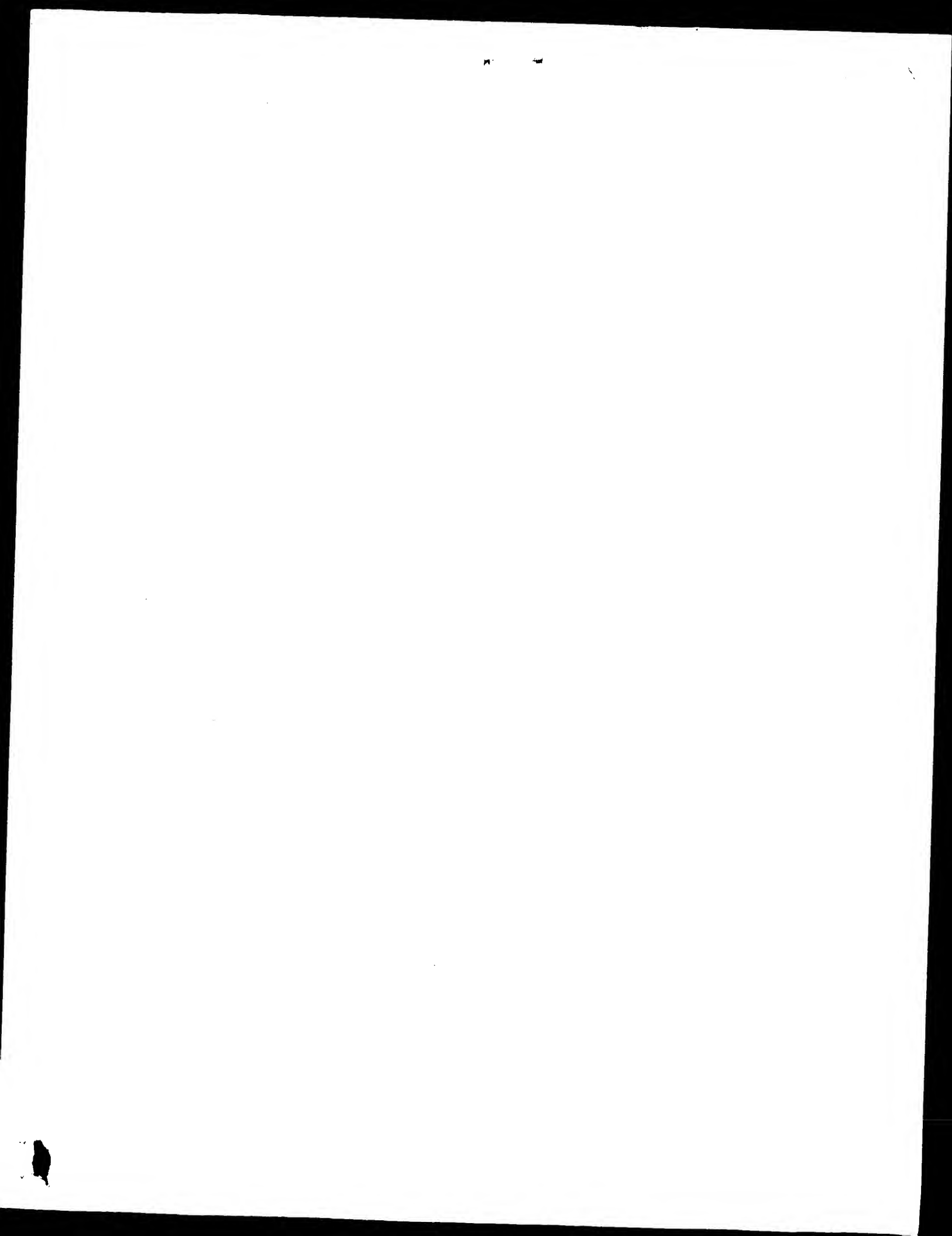
34	36	41.4	198	1 YAEI_HAEIN	P46492 haemophilus
35	36	41.4	269	1 YOEI_ECOLI	Q46942 escherichia
36	36	41.4	278	1 HT22_MYCPN	P46604 arabidopsis
37	36	41.4	312	1 LDH_MYCPN	P78007 mycoplasma
38	36	41.4	393	1 TRUA_MOUSE	Q9WU56 mus musculus
39	36	41.4	444	1 SLAP_LACAC	P35829 lactobacill
40	36	41.4	530	1 TP6B_SULSH	O05207 sulfolobus
41	36	41.4	576	1 Y134_HUMAN	Q14147 homo sapien
42	36	41.4	681	1 CRV1_ARATH	Q43135 arabidopsis
43	36	41.4	808	1 VGLB_HSVSA	P24905 herpesvirus
44	36	41.4	830	1 LEM3_HUMAN	P16109 homo sapien
45	36	41.4	925	1 PIP1_YEAST	P40020 saccharomyc

ALIGNMENTS

RESULT 1	FMA_SALTI	STANDARD;	PRT;	184 AA.
ID	FMA_SALTI			
AC	P37920;			
DT	01-OCT-1994 (Rel. 30, Created)			
DT	01-OCT-1994 (Rel. 30, Last sequence update)			
DT	01-NOV-1995 (Rel. 32, Last annotation update)			
DE	TYPE-1 FIMBRIAL PROTEIN, A CHAIN PRECURSOR (TYPE-1A PILIN).			
GN	FIMA.			
OS	Salmonella typhi.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Salmonella.			
OX	NCBI_TaxID=601;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-STY4;			
RX	MEDLINE=94116931; PubMed=8288102;			
RA	Rossolini G.M.; Muscas P.; Chiesurin A.; Satta G.;			
RT	"Analysis of the Salmonella fim gene cluster: identification of a new			
RT	gene (fimi) encoding a fimbria-like protein and located downstream			
RT	from the fima gene".			
RT	FEWS Microbiol. Lett. 114:259-266(1993);			
CC	FUNCTION: FIMBRIAL PROTEIN, A CHAIN. POLAR FILAMENTS RADIATING			
CC	FROM THE SURFACE OF THE BACTERIUM TO A LENGTH OF 0.5-1.5			
CC	MICROMETERS AND NUMBERING 100-300 PER CELL. ENABLE BACTERIA TO			
CC	COLONIZE THE EPITHELIUM OF SPECIFIC HOST ORGANS.			
CC	-1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIA PROTEINS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	EMBL; X65168; CAA46286.1; -			
DR	PIR; S20682; S20682.			
DR	InterPro; IPR000259; -			
DR	Pfam; PF00419; Fimbrin; 1.			
KW	Fimbria; Signal.			
FT	SIGNAL 1 22			
FT	CHAIN 23 184			
FT	DISULFID 46 86			
SQ	SEQUENCE 184 AA; 18793 MW; E46BADIA2A2468EB0 CRC64;			

Query Match 100.0%; Score 87; DB 1; Length 184;  
Best Local Similarity 100.0%; Pred. No. 3.6e-08; Indels 0; Gaps 0;  
Matches 18; Conservative 0; Mismatches 0;

QY 1 ASFTAIGDTTAQVFFSIV 18  
Db 65 ASFTAIGDTTAQVFFSIV 82



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 13:31:22 ; Search time 8.38 Seconds

(without alignments)

73.560 Million cell updates/sec

Title: US-09-646-043-1

Sequence: 1 ASPTAIGDTTAQVPSIV 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87	100.0	184	1	FMIA_SALTI
2	82	94.3	185	1	FMIA_SALTI
3	82	94.3	185	1	FMIA_SALTI
4	49	56.3	191	1	SEMA_ECOLI
5	47	54.0	171	1	SEMA_ECOLI
6	43	49.4	216	1	SEMA_ECOLI
7	42	48.3	205	1	SEMA_ECOLI
8	41	47.1	178	1	SEMA_ECOLI
9	41	47.1	211	1	SEMA_ECOLI
10	41	47.1	215	1	SEMA_ECOLI
11	41	47.1	215	1	SEMA_ECOLI
12	40	46.0	176	1	SEMA_ECOLI
13	40	46.0	2628	1	SEMA_ECOLI
14	39	44.8	488	1	SEMA_ECOLI
15	39	44.8	488	1	SEMA_ECOLI
16	38	43.7	170	1	SEMA_ECOLI
17	38	43.7	175	1	SEMA_ECOLI
18	38	43.7	193	1	SEMA_ECOLI
19	38	43.7	193	1	SEMA_ECOLI
20	38	43.7	193	1	SEMA_ECOLI
21	38	43.7	193	1	SEMA_ECOLI
22	37	42.5	227	1	SEMA_ECOLI
23	37	42.5	285	1	SEMA_ECOLI
24	37	42.5	354	1	SEMA_ECOLI
25	37	42.5	354	1	SEMA_ECOLI
26	37	42.5	851	1	SEMA_ECOLI
27	37	42.5	902	1	SEMA_ECOLI
28	37	42.5	1499	1	SEMA_ECOLI
29	37	42.5	1876	1	SEMA_ECOLI
30	37	42.5	1895	1	SEMA_ECOLI
31	36	41.4	132	1	SEMA_ECOLI
32	36	41.4	174	1	SEMA_ECOLI
33	36	41.4	194	1	SEMA_ECOLI

34	36	41.4	198	1	YAEI_HAEIN
35	36	41.4	269	1	YOEI_ECOLI
36	36	41.4	278	1	HT22_ARATH
37	36	41.4	312	1	LDH_MYCN
38	36	41.4	393	1	TRUA_MOUSE
39	36	41.4	444	1	SLAP_LACAC
40	36	41.4	530	1	TP6B_SULSH
41	36	41.4	576	1	Y134_HUMAN
42	36	41.4	681	1	CRY1_ARATH
43	36	41.4	809	1	VLGB_HSVSA
44	36	41.4	830	1	LEM3_HUMAN
45	36	41.4	925	1	PIPI_YEAST

## ALIGNMENTS

RESULT 1	FMIA_SALTI	STANDARD:	PRT: 184 AA.
AC	P37920:		
DR	01-OCT-1994 (Rel. 30, Created)		
DT	01-OCT-1994 (Rel. 30, Last sequence update)		
DE	TYPE-1 FIMBRIAL PROTEIN, A CHAIN PRECURSOR (TYPE-1A PILIN).		
GN	FMIA.		
OS	Salmonella typhi.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
CC	Salmonella.		
OX	NCBI_Taxid=601:		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-STY4:		
RA	MEDLINE-94116831; PubMed-8288102;		
RT	Rossolini G.M., Muscas P., Chiesurin A., Satta G.;		
RT	"Analysis of the salmonella film gene cluster: identification of a new		
RT	gene (film) encoding a fimbria-like protein and located downstream		
RT	from the film gene."		
RL	FEBS Microbiol. Lett. 114:259-266(1993)		
CC	FUNCTION: FIMBRIN (ALSO CALLED PILIN). POLAR FILAMENTS RADIATING		
CC	FROM THE SURFACE OF THE BACTERIUM TO A LENGTH OF 0.5-1.5		
CC	MICROMETERS AND NUMBERING 100-300 PER CELL. ENABLE BACTERIA TO		
CC	COLONIZE THE EPITHELIUM OF SPECIFIC HOST ORGANS.		
CC	-1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIN PROTEINS.		
CC	-1- This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).		
DR	EMBL; X65168; CAA46286.1; -		
DR	PIR; S20682; S20682.		
DR	InterPro: IPR000259; -		
DR	Pfam: PF00419; Fimbrin; 1.		
KW	Fimbria; Signal.		
FT	STGNL 1 22		POTENTIAL.
FT	CHAIN 23 184		TYPE-1 FIMBRIAL PROTEIN, A CHAIN.
FT	DISULFID 46 86		PROBABLE.
FT	SEQUENCE 184 AA; 18793 MW; E46BADI22468EB0 CMC64;		

Query Match 100.0%; Score 87; DB 1; Length 184;

Best local Similarity 100.0%; Pred. No. 3.6e-08;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 ASPTAIGDTTAQVPSIV 18

65 ASPTAIGDTTAQVPSIV 82

```

RESULT 2
FM1_SALTY STANDARD; PRT; 185 AA.
AC P55223;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE FIMBRIAL SUBUNIT TYPE 1 PRECURSOR.
OS Salmonella typhimurium.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
ON NCBI_TaxID=602;
RX SEQUENCE FROM N.A.
MEDLINE=8058806; PubMed=2890624;
RA Purcell B.K., Bruckner J., Clegg S.;
RT "Nucleotide sequences of the genes encoding type 1 fimbrial subunits
of Klebsiella pneumoniae and Salmonella typhimurium."
RL J. Bacteriol. 169:5831-5834(1987).
CC -1- SUBCELLULAR LOCATION: FIMBRIA.
CC -1- SIMILARITY: WITH E.COLI AND K.PNEUMONIAE FIMBRIAL SUBUNITS TYPE 1
CC PRECURSORS.
CC -1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIA PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; M18283; AAA27063.1; -
DR InterPro; IPR000259; -
DR Pfam; PF00419; Fimbrial; 1.
KW Fimbria; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 185 FIMBRIAL SUBUNIT TYPE 1.
FT DISULFID 46 86 PROBABLE.
SQ SEQUENCE 185 AA; 18897 MW; 52B73721952CB793 CRC64;

Query Match 94.3%; Score 82; DB 1; Length 185;
Best Local Similarity 94.4%; Pred. No. 2.7e-07;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFTAGDTTAQVPSIV 18
DB 65 ASFTAGDTTAQVPSIV 82

RESULT 3
FM1_SALTY STANDARD; PRT; 185 AA.
AC P37921;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE TYPE-1 FIMBRIAL PROTEIN, A CHAIN PRECURSOR (TYPE-1A PILIN).
OS Salmonella typhimurium.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
ON NCBI_TaxID=602;
RX SEQUENCE FROM N.A.
RA Swenson D.L., Clegg S.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
SQ SEQUENCE OF 23-50.
STRAIN=LT2 / SH6749;
aaalen K., Stetten K., Froeholm L.O., Vaeisaenen V., Korhonen T.K.;

```

```

RT "The N-terminal amino acid sequence of type 1 fimbria (pil1) of
RT Salmonella typhimurium LT2."
RL FEMS Microbiol. Lett. 16:149-151(1983).
CC -1- FUNCTION: FIMBRIAE (ALSO CALLED PILI), POLAR FILAMENTS RADIATING
CC FROM THE SURFACE OF THE BACTERIUM TO A LENGTH OF 0.5-1.5
CC MICROMETERS AND NUMBERING 100-300 PER CELL, ENABLE BACTERIA TO
CC COLONIZE THE EPITHELIUM OF SPECIFIC HOST ORGANS.
CC -1- SUBCELLULAR LOCATION: FIMBRIA.
CC -1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIA PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; L19338; AAA75416.1; -
DR StyGene; SG10275; fima.
DR InterPro; IPR000259; -
DR Pfam; PF00419; Fimbrial; 1.
KW Fimbria; Signal.
FT SIGNAL 1 22 TYPE-1 FIMBRIAL PROTEIN, A CHAIN.
FT CHAIN 23 185 PROBABLE.
FT DISULFID 46 86
SQ SEQUENCE 185 AA; 18897 MW; 7C9C6E926AD11EB2 CRC64;

Query Match 94.3%; Score 82; DB 1; Length 185;
Best Local Similarity 94.4%; Pred. No. 2.7e-07;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFTAGDTTAQVPSIV 18
DB 65 ASFTAGDTTAQVPSIV 82

RESULT 4
SFMA_ECOLI STANDARD; PRT; 191 AA.
ID SFMA_ECOLI
AC P77660;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SFM FIMBRIAL PROTEIN, A CHAIN PRECURSOR (TYPE-1A PILIN).
GN SFMA.
OS Escherichia coli.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
ON NCBI_TaxID=562;
RX SEQUENCE FROM N.A.
RA STRAIN=RI2 / M61655;
RA MEDLINE=97426517; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick J.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RL "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RX SEQUENCE FROM N.A.
RA Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA Duncan M., Federspiel N., Hyman R., Kallman S., Komp C., Kurd O.,
RA Lew H., Lin D., Mamath A., Oefner P., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: FIMBRIAE (ALSO CALLED PILI), POLAR FILAMENTS RADIATING
CC FROM THE SURFACE OF THE BACTERIUM TO A LENGTH OF 0.5-1.5
CC MICROMETERS AND NUMBERING 100-300 PER CELL, ENABLE BACTERIA TO
CC COLONIZE THE EPITHELIUM OF SPECIFIC HOST ORGANS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: FIMBRIA.
CC -1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIA PROTEINS.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE000159; AAC73632.1; -  
 DR EMBL; U82664; AAB40283.1; -  
 DR Ecogene; EG13881; sfma.  
 DR InterPro: IPR000259; -  
 DR Pfam; PF00419; Fimbrilal; 1.  
 DR Fimbrilal; Signal; 1.  
 FT SIGNAL 1 22  
 FT CHAIN 23 191  
 FT DISULFID 52 92  
 FT SEQUENCE 191 AA; 19754 MW; EBCA2BE6FE56AF3F CRC64;  
 SO  
 Query Match 56.3%; Score 49; DB 1; Length 191;  
 Best Local Similarity 53.3%; Pred. No. 0.18;  
 Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 OY 3 FTALGDTTAAOVFSI 17  
 Db 73 FNAVGTSLIRPFI 87  
 RESULT 5  
 SFME\_ECOLI STANDARD; PRT; 171 AA.  
 ID P38052; P75716; P77079;  
 AC P38052; P75716; P77079; Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE FIMBRIAL-LIKE PROTEIN SFME PRECURSOR.  
 GN SFME.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Escherichia.  
 CC NCBI\_Taxid=562;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-K12 / MG1655;  
 RC MEDLINE=9742617; PubMed=9278503;  
 RA Blatner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12."  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RP Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,  
 RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H.,  
 RA Lin D., Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 119-171 FROM N.A.  
 RC STRAIN-K12;  
 RC MEDLINE=90220507; PubMed=2183007;  
 RA Muramatsu S., Mizuno T.;  
 RT "Nucleotide sequence of the region encompassing the int gene of a  
 RT cryptic prophage and the dna y gene flanked by a curved DNA sequence  
 RT of Escherichia coli K12."  
 RL Mol. Gen. Genet. 220:325-328(1990).  
 RN [4]  
 RP IDENTIFICATION.  
 RP MEDLINE=95075659; PubMed=7984428;  
 RA Borodovsky M., Rudd K.E., Koonin E.V.;  
 RT "Intrinsic and extrinsic approaches for detecting genes in a

RT bacterial genome." 22:4756-4767(1994).  
 CC Nucleic Acids Res. 22:4756-4767(1994).  
 CC -1- SIMILARITY: BELONGS TO THE FIMB/PAPA FAMILY OF FIMBRIAL PROTEINS.  
 CC -1- SIMILARITY: STRONG, TO S.TYPHIMORIUM FIMF.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE000159; AAC73632.1; -  
 DR EMBL; U82598; AAB40732.1; ALT INIT.  
 DR EMBL; X51662; -; NOT\_ANNOTATED\_CDS.  
 DR Ecogene; EG12388; sfmf.  
 DR InterPro: IPR000259; -  
 DR Pfam; PF00419; Fimbrilal; 1.  
 DR Fimbrilal; Signal; 1.  
 FT SIGNAL 1 20  
 FT CHAIN 21 171  
 FT DISULFID 40 80  
 FT SEQUENCE 171 AA; 18222 MW; 59230981242A1DCA CRC64;  
 SO  
 Query Match 54.0%; Score 47; DB 1; Length 171;  
 Best Local Similarity 81.8%; Pred. No. 0.35;  
 Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 OY 7 GDTTAAOVFSI 17  
 Db 65 GDTTAAOVFSI 75  
 RESULT 6  
 HEDL\_HAEIN STANDARD; PRT; 216 AA.  
 ID HEDL\_HAEIN  
 AC P45992;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE MINOR FIMBRIAL SUBUNIT HIFD PRECURSOR.  
 GN HIFD.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 CC Haemophilus.  
 CC NCBI\_Taxid=727;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-EAGAN / SEROTYPE B;  
 RC MEDLINE=95012708; PubMed=7927773;  
 RA McCrea K.W., Watson W.J., Glasner J.R., Mairs C.F.;  
 RT "Identification of hifd and hife in the plus gene cluster of  
 RT Haemophilus influenzae type b strain Eagan."  
 RL Infect. Immun. 62:4922-4928(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN-AM30 (770235) / SEROTYPE B;  
 RC MEDLINE=95089703; PubMed=7997179;  
 RA van Ham M.S., van Alphen L., Mooi F.R., van Putten J.P.M.;  
 RT "The fimbrial gene cluster of Haemophilus influenzae type b."  
 RL Mol. Microbiol. 13:673-684(1994).  
 CC -1- FUNCTION: MAY BE A MINOR STRUCTURAL PROTEIN REQUIRED FOR PLUS  
 CC BIOGENESIS.  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID  
 CC ANCHOR (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE FIMB/PAPA FAMILY OF FIMBRIAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

Query Match	49.48;	Score 43;	DB 1;	Length 216
Best Local Similarity	52.98;	Pred. No. 2.3;		
Matches	9;	Conservative	2;	Matches

[illegible]

RESULT 7	
HF22_HAEIN	
ID HF22_HAEIN	STANDARD;
AC P45993;	PRT; 205 AA

01-NOV-1995 (Rel. 32, Created)  
01-NOV-1995 (Rel. 32, Last sequence update)  
01-NOV-1995 (Rel. 32, Last annotation update)  
MINOR FIBRILLAR SUBUNIT HIFD PRECURSOR.  
HIFD.

Haemophilus influenzae;  
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
Haemophilus;  
NCBI\_TaxID=727;  
(1)

SEQUENCE FROM N.A.  
SFRAIN=86-0295 / LKP SEROTYPE 1;  
Green B.A., Olmsted S.B.;  
Submitted (JAN-1995) to the FWRB

- FUNCTION: MAY BE A MINOR STRUCTURAL PROTEIN REQUIRED FOR PILUS BIOGENESIS.
- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID ANCHOR (PROBABLE).
- STRAIN(S)

-----  
 CATALANITY: BELONGS TO THE FIMA/PAPA FAMILY OF FLMBRIA PROTEINS.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collabbe  
 between the Swiss Institute of Bioinformatics and the EMBL outstaites  
 the European Bioinformatics Institute.  
 -----

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce>, or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL, U19730; AAA61817.1, -,  
InterPro, IPR000259, -,  
InterPro, IPR000437, -,  
Pfam, PF00449, Fimbrin, 1

PROSITE; PS00013; PROKAR LIPOPROTEIN; FALSE\_NEG  
Fimbrlia; Signal; Lipoprotein.  
SIGNAL 1 19 PROBABLE.  
CHAIN 20 205 MINOR FIMBRIAL SUBU  
LIPID 20 20

SEQUENCE	205 AA;	21663 MW;	0DA047DFAFABCDDB5 CRC64;	N-ACYL DIGLYCERIDE (PROBABLE)
Query Match	48.3%	Score 43.0	PF 1	

Best Local Matches	Similarity	Score	DB	Length
9; Conservative	52.9%	42;	1;	205;
		Pred. No. 3.2;		
		2; Mismatches	6; Indels	0; Gaps
				0

```
QY      1 ASFTAGDTTAQVPFSI 17
        | : : | | | | |
Db      96 ANLSHAGQTAAVPFSI 112
```

RESULT	8	
LPFA_SALTY		
ID	LPFA_SALTY	
AC	P43660;	STANDARD;
		PRT; 178 AA

01-NOV-1995 (Rel. 32, Created)  
01-NOV-1995 (Rel. 32, last sequence update)  
01-NOV-1995 (Rel. 32, last annotation update)  
LONG POLAR FIBRIN A PRECURSOR.

05 *Salmonella typhimurium*.  
06 Bacteria; Proteobacteria, gamma subdivision; Enterobacteriaceae  
07 *Salmonella*.  
08 NCBI\_TaxID=602;  
09

SEQUENCE FROM N.A.  
STRAIN=ATCC 14028;  
MEDLINE=95238281; PubMed=7721701

Identification and sequence analysis of *lpfABCE*, a putative fimbrial operon of *Salmonella typhimurium*.<sup>8</sup> J. Bacteriol. 177:2087-2097(1995).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on use by any user.

modified and this statement is not removed. Usage by and for commercial  
entitities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

R 100359; AAA/3966.L;  
 R StyGene; SG10557; 1pfa.  
 R InterPro; IPR000259; -;  
 R Pfam; PF00419; Fimbrial; 1.  
 R Fimbrin; Strep1

SIGNAL	1	24	POTENTIAL.
CHAIN	25	178	LONG POLAR FIBRIN PROTEIN A.
SEQUENCE	178 AA;	18683 MM;	4F4E562078B928BD CRC64;

Query Match	47.18;	Score 41;	DB 1;	Length 178;
Best Local Similarity	53.3%;	Pred. No. 4.2;		
Matches	8;	Conservative	2;	Mismatches 5;
				Indels

```

3  FTAIGDITLAVPPSI 17
   | | | | : | | |
64 FKAIGDKSSSKPRQI 78

```

SULT	9	
V_BPP2		
VPV_BPP2		STANDARD;
P31340;		PRT; 211 AA

01-JUL-1993 (Rel. 26, Created)  
01-OCT-1996 (Rel. 34, Last sequence update)  
30-MAY-2000 (Rel. 39, Last annotation update,  
BASEPLATE ASSEMBLY PROTEIN V (GPV)).

Bacteriophage P2.  
Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae.  
P2-like viruses.  
NCBI TaxID=10670.

[1] SEQUENCE FROM N.A.  
MEDLINE=96036485; PubMed=7483254;  
Haggarud-Ljungquist E, Jacobsen P, Riebow S  
et al. 1996



Thu Jun 21 08:38:34 2001

us-09-646-043-1.rsp

Calendar R.; genes involved in baseplate assembly."  
RT "Bacteriophage P2: genes involved in baseplate assembly."  
RL Virology 213:109-121(1995).  
RN (2)  
RP SEQUENCE OF 1-72 FROM N.A.  
RX MEDLINE-8516085; PubMed-3981640;  
RA "Bacteriophage P2 late promoters. II. Comparison of the four late  
RT promoter sequences." J. Mol. Biol. 181:373-382(1985).  
RL (3)  
RN REVISIONS, SEQUENCE FROM N.A.  
RP Christie G.E.;  
RA submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.  
RL (4)  
RN SEQUENCE OF 1-24 FROM N.A.  
RP MEDLINE-9423369; PubMed-8178426;  
RX Linderoth N.A., Julien B., Flick K.E., Calendar R., Christie G.E.;  
RA "Molecular cloning and characterization of bacteriophage P2 genes R  
RT and S involved in tail completion." Virology 200:347-359(1994).  
RL Virology 200:347-359(1994).  
CC -1- FUNCTION: FORMS THE SMALL SPIKE AT THE TIP OF THE TAIL.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC EMBL: AF063097; AAD03282.1;  
DR SEQUENCE 211 AA; 22244 MW; B1CA7E47B19B91FB CRC64;  
SQ  
Query Match 47.1%; Score 41; DB 1; Length 211;  
Best Local Similarity 53.3%; Pred. No. 5;  
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;  
OY 1 ASPTAGDTTAAQVPSFV 18  
DB 127 ASVTAGSSTRTVTVVWV 144  
RESULT 10  
FIM1\_ECOLI STANDARD; PRT; 215 AA.  
ID FIM1\_ECOLI  
AC P39264;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE FIMBRIN-LIKE PROTEIN FIM1.  
GN FIM1.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OC NCB1\_TaxID=562;  
OX (1)  
RN SEQUENCE FROM N.A.  
RP STRAIN-K12 / MG1655;  
RX MEDLINE-95334362; PubMed-7610040;  
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,  
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the  
RT region from 92.8 through 100 minutes."  
RL Nucleic Acids Res. 23:2105-2119(1995).  
RN (2)  
RP PRELIMINARY SEQUENCE OF 1-116 FROM N.A.  
RX Klemm P.;  
RA "The fima gene encoding the type-1 fimbrial subunit of Escherichia  
RT coli. Nucleotide sequence and primary structure of the protein."  
RL Eur. J. Biochem. 143:395-399(1984).

CC -1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIA PROTEINS.  
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO  
CC FRAMESHIFTS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC EMBL: U14003; AAA97211.1;  
DR EMBL: AE005002; AAC77211.1;  
DR EMBL: X00981; NOT\_ANNOTATED\_CDS.  
DR EcoGene: EG11974; FIM1.  
DR InterPro: IPR000259;  
DR Pfam: PF00419; Fimbrin1.1.  
DR Fimbrin.  
KM DISULFID 79 119 BY SIMILARITY.  
FT SEQUENCE 215 AA; 23583 MW; 601F727DE4864B3 CRC64;  
SQ  
Query Match 47.1%; Score 41; DB 1; Length 215;  
Best Local Similarity 53.3%; Pred. No. 5; Indels 0; Gaps 0;  
Matches 8; Conservative 3; Mismatches 4;  
OY 3 FRAIGDTTAAQVPSFV 17  
DB 100 FHAVGDSAPVPEVI 114  
RESULT 11  
V11\_HPV31 STANDARD; PRT; 504 AA.  
ID V11\_HPV31  
AC P17388;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE MAJOR CAPSID PROTEIN L1.  
GN L1.  
OS Human papillomavirus type 31.  
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.  
OC NCB1\_TaxID=10585;  
OX (1)  
RN SEQUENCE FROM N.A.  
RP MEDLINE-89299478; PubMed-2545036;  
RX Goldborough M.D., Dislyestre D., Temple G.F., Lorincz A.T.;  
RA "Nucleotide sequence of human papillomavirus type 31: a cervical  
RT neoplasia-associated virus." Virology 171:306-311(1989).  
RL Virology 171:306-311(1989).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC EMBL: J04353; AAA6956.1;  
DR PIR: G32444; P1WL31.  
DR InterPro: IPR002210;  
DR Pfam: PF00500; Late\_protein\_L1.1.  
DR PRINTS: PR000865; HPVCAPSIDL1.  
KM Coat protein; Late protein.  
FT SEQUENCE 504 AA; 56352 MW; B45A306A6B3AB9D2 CRC64;  
SQ  
Query Match 47.1%; Score 41; DB 1; Length 504;  
Best Local Similarity 53.3%; Pred. No. 13;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 FTAIGDTTAAQVPSI 17  
 DB 211 FTAIGDTTAAQVPSI 225

## RESULT 12

FT FIMB\_ECOLI STANDARD: PRT: 176 AA.  
 AC P08189;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DE FIMB PROTEIN PRECURSOR.  
 GN FIMB  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 ON Escherichia.  
 OX NCBI\_Taxid=562;  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=88038337; PubMed=2890081;  
 RT Klemm P., Christiansen G.;  
 RT "Three fimb genes required for the regulation of length and mediation  
 of adhesion of Escherichia coli type 1 fimbriae.";  
 RL Mol. Gen. Genet. 208:439-445(1987).  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RA MEDLINE=9534362; PubMed=7610040;  
 RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,  
 RA Blattner F.R.;  
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the  
 region from 92.8 through 100 minutes.";  
 RL Nucleic Acids Res. 23:2105-2119(1995).  
 RP SEQUENCE OF 1-3 FROM N.A.  
 RC STRAIN-K12;  
 RA MEDLINE=9020509; PubMed=1970114;  
 RA Klemm P., Christiansen G.;  
 RT "The fimb gene required for cell surface localization of Escherichia  
 coli type 1 fimbriae.";  
 RL Mol. Gen. Genet. 220:334-338(1990).  
 RP FIMBRIAE INVOLVED IN REGULATION OF LENGTH AND MEDIATION OF  
 ADHESION OF TYPE 1 FIMBRIAE (BUT NOT NECESSARY FOR THE PRODUCTION  
 OF FIMBRIAE). INVOLVED IN THE INTEGRATION OF FIMB IN THE FIMBRIAE.  
 CC -1- SIMILARITY: TO TYPE-1 FIMBRIAL SUBUNITS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL: X05672; CA29154.1;  
 DR EMBL: U14003; AA97214.1;  
 DR EMBL: AE000502; AAC7274.1;  
 DR EMBL: X51655; CA35969.1;  
 DR PTR: S07321; S07321.  
 DR Ecogene: EG10313; FIMF.  
 DR InterPro: IPR000259;  
 DR Pfam: PF00419; Fimbrin; 1.  
 DR Fimbrin; Signal.  
 FT CHAIN 1 20  
 FT DISULFID 21 176  
 FT SITE 38 78  
 FT 175 175  
 FT POTENTIAL.  
 FT FIMB PROTEIN.  
 FT PROBABLE.  
 FT REQUIRED FOR STABILITY AND TRANSPORT (BY  
 FT SIMILARITY)  
 FT P -> S (IN REF. 1).  
 FT S -> L (IN REF. 1).  
 FT A -> V (IN REF. 1).  
 FT AVKGTGV -> RRGVYWR (IN REF. 1).  
 FT CONFLICT 70 70  
 FT CONFLICT 76 76  
 FT CONFLICT 81 81  
 FT CONFLICT 84 92

FT CONFLICT 107 107  
 SQ SEQUENCE 176 AA; 18715 MW; V -> A (IN REF. 1).  
 38692EE6A40121P CRC64;

## Query Match

Best Local Similarity 46.0%; Score 40; DB 1; Length 176;  
 Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
 OY 3 FTAIGDTTAAQVPSI 18  
 DB 59 FNNIGATPEVPEFRIL 74

## RESULT 13

FT HAGA\_PORGI STANDARD: PRT: 2628 AA.  
 AC Q51845;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DE HEMAGGLUTININ A PRECURSOR.  
 GN HAGA  
 OS Porphyromonas gingivalis (Bacteroides gingivalis).  
 OC Bacteria; CF8 group; Bacteroidaceae; Porphyromonas.  
 ON Bacteroides.  
 OX NCBI\_Taxid=837;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-381;  
 RA MEDLINE=9704672; PubMed=8926061;  
 RA Han N., Whitlock J., Prognulski-Fox A.;  
 RT "The hemagglutinin gene A (haga) of Porphyromonas gingivalis 381  
 contains four large, contiguous, direct repeats.";  
 RL Infect. Immun. 64:4000-4007(1996).  
 CC -1- FUNCTION: AGGLUTININATES ERYTHROCYTES.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C25.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL: U41807; AAB17128.1;  
 DR InterPro: IPR001769;  
 DR Pfam: PF01364; Peptidase\_C25; 5.  
 DR Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat.  
 FT CHAIN 1 24  
 FT DOMAIN 25 2628  
 FT DOMAIN 25 539  
 FT DOMAIN 540 995  
 FT DOMAIN 996 1451  
 FT DOMAIN 1452 1907  
 FT DOMAIN 2074 2628  
 FT SEQUENCE 2628 AA; 283324 MW; 61C4DE32540C99DA CRC64;  
 Query Match  
 Best Local Similarity 46.0%; Score 40; DB 1; Length 2628;  
 Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;  
 OY 2 SFTAIGDTTAAQVPSI 17  
 DB 2554 NITSADVTAQKPYTL 2569

Thu Jun 21 08:38:34 2001

us-09-646-043-1.rsp

Page 7

01-OCT-2000 (rel. 40, last annotation update)  
DE ALKALINE EXONUCLEASE (EC 3.1.11.-).  
GN U70 OR 16R.  
OS Human herpesvirus (type 6 / strain Uganda-1102) (HHV6).  
CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
CC Betaherpesvirinae; Roseolovirus.  
OX NCBI\_TaxID=10370;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90080132; PubMed=2152817;  
RA Lawrence G.L., Chee M., Craxton M.A., Honess R.W.,  
RT J. Virol. 64:287-299(1990).  
RL "Human herpesvirus 6 is closely related to human cytomegalovirus."  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95266321; PubMed=7747482;  
RA Compels U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,  
RT Martin M.E., Etstathiou S., Craxton M., Macaulay H.A.;  
RL "The DNA sequence of human herpesvirus-6: structure, coding content,  
RT and genome evolution."  
CC VIROLOGY 209:29-51(1995).  
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES ALKALINE EXONUCLEASE  
CC FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL: X83413; CAA5362.1; -  
CC DR EMBL: M68963; AAB5578.1; -  
CC DR PIR: F36769; GOBHS.  
CC DR InterPro: IPR001616; -  
CC DR Pfam: PF01771; Herpes\_alk\_exo; 1.  
CC DR PRINTS: PR00924; ALKEXNUCLASE.  
CC DR HydroLase: Nuclease; Exonuclease.  
KW KW HydroLase: Nuclease; Exonuclease.  
SQ SEQUENCE 488 AA; 56644 MW; 0F38A10597366A5B CRC64;  
  
Query Match 44.8%; Score 39; DB 1; Length 488;  
Best Local Similarity 64.3%; Pred. NO. 28;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
  
OY 5 AIGDTAQPFSIV 18  
DB 29 AIREKIKQVPSIV 42  
II : |||||  
  
RESULT 15  
EXON\_HSV62 STANDARD; PRT; 488 AA.  
ID EXON\_HSV62  
AC P52448;  
DT 01-OCT-1996 (rel. 34, Created)  
DT 01-OCT-1996 (rel. 34, Last sequence update)  
DT 01-OCT-2000 (rel. 40, Last annotation update)  
DE ALKALINE EXONUCLEASE (EC 3.1.11.-).  
GN U70 OR CH3R.  
OS Human herpesvirus (type 6 / strain 229) (HHV6).  
CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
CC Betaherpesvirinae; Roseolovirus.  
OX NCBI\_TaxID=16351;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=8634027; PubMed=8634027;  
RA Lindquester G.J., Inoue N., Allen R.D., Castelli J.W.,  
RA Stamey F.R., Dambaugh T.R., O'Brian J.J., Danovich R.M.,  
RA Frenkel N., Pellett P.E.,  
RT "Restriction endonuclease mapping and molecular cloning of the human  
herpesvirus 6 variant B strain 229 genome."  
RT

Arch. Virol. 141:367-379(1996).  
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES ALKALINE EXONUCLEASE  
CC FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL: AF157706; AAB06353.1; -  
CC DR InterPro: IPR001616; -  
CC DR Pfam: PF01771; Herpes\_alk\_exo; 1.  
CC DR PRINTS: PR00924; ALKEXNUCLASE.  
CC DR HydroLase: Nuclease; Exonuclease.  
KW KW HydroLase: Nuclease; Exonuclease.  
SQ SEQUENCE 488 AA; 56687 MW; AE2872028D4B3D90 CRC64;  
  
Query Match 44.8%; Score 39; DB 1; Length 488;  
Best Local Similarity 64.3%; Pred. NO. 28;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
  
OY 5 AIGDTAQPFSIV 18  
DB 29 AIREKIKQVPSIV 42  
II : |||||  
  
Search completed: June 20, 2001, 13:32:54  
Job time: 92 sec

Thu Jun 21 08:38:34 2001

us-09-646-043-1.rsp

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using SW model

Run on: June 20, 2001, 13:31:22 ; Search time 13.45 Seconds

(without alignments)  
101.944 Million cell updates/sec

Title: US-09-646-043-1

Sequence: 1 ASFTAGDTPAQVPFSIV 18

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	100.0	184	2	S20682 type 1 fimbrial pr
2	82	94.3	185	2	B28393 type 1 fimbrial pr
3	49	56.3	191	2	A64785 type 1 fimbrial pr
4	49	56.3	191	2	C85553 probable fimbrial-
5	47	54.0	171	2	B64785 fimbrial protein h
6	44.5	51.1	591	2	S65588 ABC-transporter st
7	44	50.6	171	2	G85553 probable fimbrial
8	44	50.6	931	2	B85062 probable WD-repeat
9	43	49.4	216	2	S54430 fimbrial protein h
10	43	49.4	659	2	T33557 hypothetical prote
11	43	49.4	1248	2	B86827 hypothetical prote
12	42	48.3	194	2	G85976 probable fimbrial-
13	42	48.3	144	2	T41416 probable tyrosine-
14	42	48.3	1146	2	E70204 hypothetical prote
15	41	47.1	177	2	A66271 long polar fimbria
16	41	47.1	199	2	A61725 conserved hypotet
17	41	47.1	215	2	S56540 fimi protein - Esc
18	41	47.1	215	2	E86129 fimbrial protein l
19	41	47.1	504	1	PIWL31 L1 protein - human
20	41	47.1	628	2	F84219 Hc16 transducer l
21	41	47.1	1298	2	T47523 DNA-binding protei
22	40	46.0	176	2	S65543 fimbrial protein f
23	40	46.0	1419	2	H86129 fimbrial morpholog
24	40	46.0	1732	2	T30836 hypothetical prote
25	40	46.0	2628	2	T28651 lysine-specific cy
26	40	46.0	270	2	T43039 hemagglutinin A -
27	39	44.8	270	2	T43039 hypothetical prote
28	39	44.8	409	2	S77262 hypothetical prote
29	39	44.8	468	2	B70932 probable PPE prote

30	39	44.8	488	1	OOBHS	alkaline exonuclease
31	39	44.8	488	2	T44030	alkaline exonuclease
32	39	44.8	488	2	T44215	alkaline exonuclease
33	39	44.8	615	2	T39758	proline-serine ric
34	39	44.8	725	2	E86790	hypothetical prote
35	38	43.7	85	2	T17696	hypothetical prote
36	38	43.7	170	2	I38892	hypothetical prote
37	38	43.7	175	2	S06193	sfag protein precu
38	38	43.7	175	2	I59446	foef protein - Esc
39	38	43.7	178	2	C86029	probable major fim
40	38	43.7	202	2	C71306	hypothetical prote
41	38	43.7	218	2	F75404	ABC transporter, p
42	38	43.7	259	2	T23782	hypothetical prote
43	38	43.7	347	2	S43771	phosphatidylcholin
44	38	43.7	458	2	T27217	hypothetical prote
45	38	43.7	460	2	A84261	hypothetical prote

## ALIGNMENTS

RESULT 1  
S20682  
type 1 fimbrial protein fima - Salmonella typhi  
C:Species: Salmonella typhi  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 26-Aug-1999  
C:Accession: S20682  
R:ROSSOLINI, G.M.; MUSCAS, P.; CHIESURIN, A.; SARTI, G.  
submitted to the EMBL Data Library, March 1992  
A:Reference number: S20682  
A:Accession: S20682  
A:Molecule type: DNA  
A:Residues: 1-184 <ROS>  
A:Cross-references: EMBL:X65168; NID:9395382; PID:CAA46286.1; PID:947667  
A:Experimental source: strain Sty4  
C:Genetics:  
A:Gene: fima  
C:Superfamily: type 1 fimbrial protein  
C:Keywords: fimbria

Query Match 100.0%; Score 87; DB 2; Length 184;  
Best Local Similarity 100.0%; Pred. No. 8.6e-08;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFTAGDTPAQVPFSIV 18  
DB 65 ASFTAGDTPAQVPFSIV 82

RESULT 2  
B28393  
type 1 fimbrial protein precursor - Salmonella typhimurium  
C:Species: Salmonella typhimurium  
C:Date: 28-Aug-1989 #sequence\_revision 28-Aug-1989 #text\_change 15-Nov-1996  
C:Accession: B28393; A05121  
R:Purcell, B.K.; Pruckler, J.; Clegg, S.  
J. Bacteriol. 169, 5831-5834, 1987  
A:Title: Nucleotide sequences of the genes encoding type 1 fimbrial subunits of Klebs  
A:Reference number: A91858; WUID:88058806  
A:Accession: B28393  
A:Molecule type: DNA  
A:Residues: 1-185 <PUR>  
A:Note: The authors translated the codon GCC for residue 18 as Gly, GNG for residue 3  
R:Maalen, K.; Sletten, K.; Froholm, L.O.; Valsanen, V.; Korhonen, T.K.  
FEMS Microbiol. Lett. 16, 149-151, 1983  
A:Reference number: A05121  
A:Accession: A05121  
A:Molecule type: protein  
A:Residues: 23-30, 'S', 32-45, 'X', 47-50 <MAA>  
C:Superfamily: type 1 fimbrial protein  
F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-185/Product: type 1 fimbrial protein #status predicted <MAT>



Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: AB5480; MUID:21074935; PMID:11206551  
 A:Accession: G85553  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-171 <STO>  
 A:Cross-references: GB:AE005174; NID:912513434; PIDN:AG54891.1; GSPDB:GN00145; UWCP:206  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: stx2f  
 C:Superfamily: type 1 fimbrial protein

Query Match  
 Best Local Similarity 50.6%; Score 44; DB 2; Length 171;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 7 GDTTAAQVPSI 17  
 |||||:  
 DB 65 GDTTALLPFSL 75

RESULT 8  
 B85062  
 probable WD-repeat membrane protein [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
 C:Accession: B85062  
 R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
 Nature 402, 769-777, 1999  
 A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
 A:Reference number: AB5001; MUID:20083488  
 A:Accession: B85062  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-931 <STO>  
 A:Cross-references: GB:NC\_001268; NID:97267253; PIDN:CAB81036.1; GSPDB:GN00140  
 A:Genetics:  
 A:Gene: AT4G04940  
 A:Map position: 4

Query Match  
 Best Local Similarity 50.6%; Score 44; DB 2; Length 931;  
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 QY 3 FTAIGDTTAAQVPSI 17  
 |||||:  
 DB 7 FTAIGTITVPSV 21

RESULT 9  
 S54430  
 fimbrial protein hlfD precursor - Haemophilus influenzae (strain AM30)  
 C:Species: Haemophilus influenzae  
 C:Date: 06-Sep-1996 #sequence\_revision 13-Mar-1997 #text\_change 24-Nov-1999  
 C:Accession: S54430  
 R:Van Ham, S.M.; Van Alphen, L.; Mooi, F.R.; van Putten, J.P.M.  
 Mol. Microbiol. 13, 673-684, 1994  
 A:Title: The fimbrial gene cluster of Haemophilus influenzae type b.  
 A:Reference number: S54428; MUID:95089703  
 A:Accession: S54430  
 A>Status: preliminary; nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 1-216 <VAN>  
 A:Cross-references: EMBL:Z53502; NID:9535165; PIDN:CAA83903.1; PID:9535168  
 C:Genetics:  
 A:Gene: hlfD  
 C:Superfamily: type 1 fimbrial protein

Query Match 49.4%; Score 43; DB 2; Length 216;

Best Local Similarity 52.9%; Pred. No. 5.2;  
 Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 ASFTAGDTTAAQVPSI 17  
 |||||:  
 DB 90 ANLSLAGTAAQVPSI 106

RESULT 10  
 T33557  
 hypothetical protein W04C9.6 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T33557  
 R:Fulton, R.; Rohlfing, T.; Morris, M.  
 submitted to the EMBL Data Library, October 1998  
 A:Description: The sequence of C. elegans cosmid W04C9.  
 A:Reference number: Z21368  
 A:Accession: T33557  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-659 <FUL>  
 A:Cross-references: EMBL:AR098999; PIDN:AC68729.1; GSPDB:GN00019; CESP:W04C9.6  
 A:Genetics:  
 A:Gene: CESP:W04C9.6  
 A:Map position: 1  
 A:Introns: 11/3; 48/1; 113/1; 240/3; 320/2; 361/3; 393/3; 417/2; 477/2

Query Match 49.4%; Score 43; DB 2; Length 659;  
 Best Local Similarity 56.2%; Pred. No. 18;  
 Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
 QY 2 SFTAIGDTTAAQVPSI 17  
 |||||:  
 DB 60 SCVAIGSLAATPSI 75

RESULT 11  
 B96827  
 hypothetical protein TBK14.1 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: B96827  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, N.F.; Hughes, B.; Huizart, L.  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallo, K.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: AB6141; MUID:21016719  
 A:Accession: B96827  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1248 <STO>  
 A:Cross-references: GB:AE005173; NID:94835752; PIDN:AAD30219.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: TBK14.1  
 A:Map position: 1

Query Match 49.4%; Score 43; DB 2; Length 1248;  
 Best Local Similarity 42.9%; Pred. No. 38;  
 Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ASFTAGDTTAAQV 14  
 |||||:  
 DB 464 SNYSAGTSSIP 477

## RESULT 12

probable fimbrial-like protein yrah [imported] - Escherichia coli (strain O157:H7)

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001

C:Accession: G85976

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoulsis, K.; Apodaca,

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: G85976

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-194 <STO>

A:Cross-references: GB:AE005174; NID:g12517744; PIDN:AA658275.1; GSPDB:GN00145; UWGP:244

C:Genetics:

A:Gene: yrah

Query Match

Best Local Similarity 48.3%; Score 42; DB 2; Length 194;

Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 3 FTAIGDTTAQVPFST 17

DB 59 FRSVODRSPKIPFTI 73

## RESULT 13

T41416

probable tyrosine-tRNA ligase precursor, mitochondrial - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000

C:Accession: T41416

R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.

submitted to the EMBL Data Library, September 1998

A:Reference number: Z21954

A:Accession: T41416

A:Status: preliminary; translated from GB/EMBL/DBDJB

A:Molecule type: DNA

A:Residues: 1-445 <WOO>

A:Cross-references: EMBL:AL031798; PIDN:CAA21185.1; GSPDB:GN00068; SPDB:SPCC576.06c

A:Experimental source: strain 972h-; cosmid c576

C:Genetics:

A:Gene: SPDB:SPCC576.06c

A:Map position: 3

A:Genome: nuclear

A:Introns: 22/2; 268/3

C:Superfamily: tyrosine-tRNA ligase

C:Keywords: mitochondrion

Query Match

Best Local Similarity 48.3%; Score 42; DB 2; Length 445;

Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 FTAIGDTTAQVPFST 13

DB 65 FTVIGDATAQOL 75

## RESULT 14

E70204

hypothetical protein B80838 - Lyme disease spirochete

C:Species: Borrelia burgdorferi (Lyme disease spirochete)

C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999

C:Accession: E70204

R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Iatigra, R.; White

son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,

; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997

A:Authors: Smith, H.O.; Venter, J.C.

A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.

A:Reference number: A70100; MUID:96065943

A:Accession: E70204

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1146 <KLE>

A:Cross-references: GB:AE001182; GB:AE000783; NID:q2688780; PIDN:AA67188.1; PID:q268

A:Experimental source: strain B31

Query Match

Best Local Similarity 48.3%; Score 42; DB 2; Length 1146;

Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 3 FTAIGDTTAQVPFST 17

DB 948 FSKIGTTTKLVPYSL 962

## RESULT 15

A56271

long polar fimbrial major protein precursor - Salmonella typhimurium

C:Species: Salmonella typhimurium

C:Date: 03-Oct-1995 #sequence\_revision 03-Oct-1995 #text\_change 07-May-1999

C:Accession: A56271

R:Baumler, A.J.; Heffron, F.

J. Bacteriol. 177, 2087-2097, 1995

A:Title: Identification and sequence analysis of lpfABCD, a putative fimbrial operon

A:Reference number: A56271; MUID:95238281

A:Accession: A56271

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-177 <BAE>

A:Cross-references: GB:U18559

C:Genetics:

A:Gene: lpfA

C:Superfamily: type 1 fimbrial protein

Query Match

Best Local Similarity 47.1%; Score 41; DB 2; Length 177;

Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 FTAIGDTTAQVPFST 17

DB 64 FKAIGDKSSSKPFQI 78

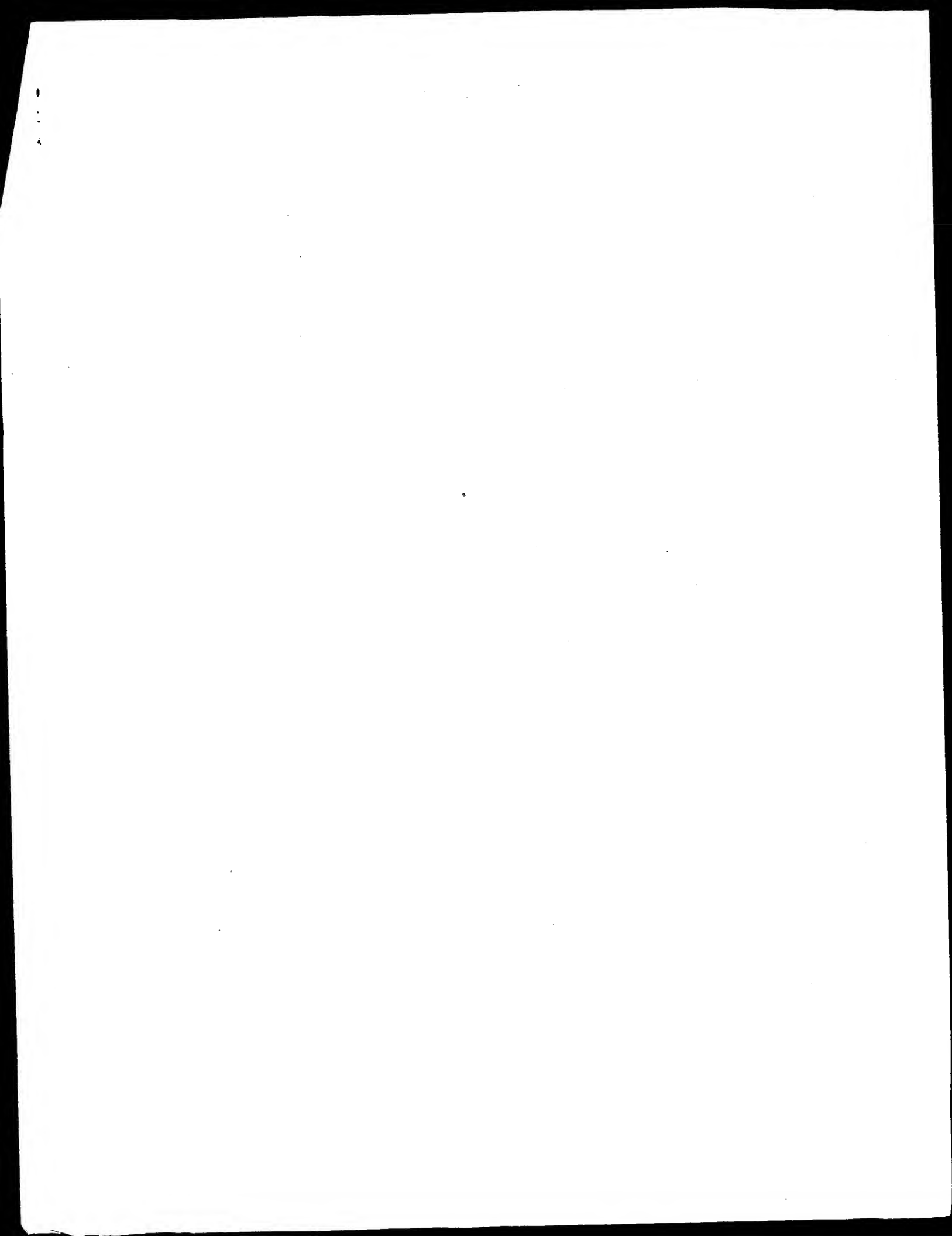
Search completed: June 20, 2001, 13:32:19

Job time: 57 sec



Jun 21 08:38:33 2001

us-09-646-043-1.rpt



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 13:31:22 ; Search time 19.45 seconds  
(without alignments)  
122.442 Million cell updates/sec

Title: US-09-646-043-1  
Perfect score: 87  
Sequence: 1 ASFTAIGDTAQPFSIV 18

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 133305027 residues  
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP:REMBL\_16:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organella:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_prodent:\*  
12: sp\_unclassified:\*  
13: sp\_vertebrate:\*  
14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	100.0	160	2 Q9X3T3	Q9X3T3 salmonella
2	84	96.6	161	2 Q9X3U2	Q9X3U2 salmonella
3	84	96.6	161	2 Q9WM13	Q9WM13 salmonella
4	82	94.3	161	2 Q9X3T5	Q9X3T5 salmonella
5	82	94.3	185	2 Q53483	Q53483 salmonella
6	79	90.8	161	2 Q9WM63	Q9WM63 salmonella
7	76	87.4	161	2 Q9X3U1	Q9X3U1 salmonella
8	76	87.4	161	2 Q9WVW8	Q9WVW8 salmonella
9	76	87.4	161	2 Q9S644	Q9S644 salmonella
10	72	82.8	161	2 Q9X3T9	Q9X3T9 salmonella
11	72	82.8	161	2 Q9X3T7	Q9X3T7 salmonella
12	72	82.8	161	2 Q9S643	Q9S643 salmonella
13	52	59.8	186	2 P72209	P72209 protease mir
14	46	52.9	250	10 Q9FXV9	Q9FXV9 glycine mtr
15	46	52.9	510	10 Q9LRW3	Q9LRW3 arabinidops
16	45	51.7	204	2 O87336	O87336 haemophilus
17	45	51.7	216	2 O86218	O86218 haemophilus
18	44.5	51.1	591	2 O54204	O54204 streptomyces
19	44	50.6	931	10 Q9ZPH3	Q9ZPH3 arabinidops

20	43	49.4	221	2 P94814	P94814 haemophilus
21	43	49.4	659	5 Q9TZE1	Q9TZE1 caenothabdi
22	43	49.4	1248	10 Q9SAJ2	Q9SAJ2 arabinidops
23	42	48.3	292	2 O85245	O85245 thermotoga
24	42	48.3	359	5 Q9GUB4	Q9GUB4 giardia lam
25	42	48.3	445	3 O74890	O74890 schistosacch
26	42	48.3	1146	2 O51778	O51778 borrelia bu
27	42	47.1	199	2 Q9PL66	Q9PL66 chlamydia m
28	41	47.1	210	2 O87338	O87338 haemophilus
29	41	47.1	504	14 Q81017	Q81017 human papil
30	41	47.1	628	1 Q9HRN6	Q9HRN6 halobacteri
31	41	47.1	1298	10 Q9LZU7	Q9LZU7 arabinidops
32	40	46.0	176	2 O87737	O87737 escherichia
33	40	46.0	176	2 Q9P626	Q9P626 escherichia
34	40	46.0	205	2 O87470	O87470 haemophilus
35	40	46.0	1097	2 P72196	P72196 porphyromon
36	40	46.0	1223	2 Q9ZNB5	Q9ZNB5 porphyromon
37	40	46.0	1419	5 O45092	O45092 caenothabdi
38	40	46.0	1732	2 O51817	O51817 porphyromon
39	40	46.0	1732	2 O52050	O52050 porphyromon
40	40	46.0	1732	2 O07442	O07442 porphyromon
41	39	44.8	201	2 Q9K3U2	Q9K3U2 streptomyces
42	39	44.8	205	2 O86217	O86217 haemophilus
43	39	44.8	358	10 Q9ZP40	Q9ZP40 plasmid saliv
44	39	44.8	409	2 P73556	P73556 synechocyst
45	39	44.8	411	5 Q9VM76	Q9VM76 drosophila

## ALIGNMENTS

Result	ID	Query Match	Score	DB 2:	Length	Pred. No.	Score	DB 2:	Length	Matches	Conservative	Mismatches	Indels	Gaps
1	Q9X3T3	100.0%	87	100.0%	160	100.0%	87	100.0%	160	18	0	0	0	0
2	Q9X3U2	96.6%	84	96.6%	161	96.6%	84	96.6%	161	18	0	0	0	0
3	Q9WM13	96.6%	84	96.6%	161	96.6%	84	96.6%	161	18	0	0	0	0
4	Q9X3T5	94.3%	82	94.3%	161	94.3%	82	94.3%	161	18	0	0	0	0
5	Q53483	94.3%	82	94.3%	185	94.3%	82	94.3%	185	18	0	0	0	0
6	Q9WM63	90.8%	79	90.8%	161	90.8%	79	90.8%	161	18	0	0	0	0
7	Q9X3U1	87.4%	76	87.4%	161	87.4%	76	87.4%	161	18	0	0	0	0
8	Q9WVW8	87.4%	76	87.4%	161	87.4%	76	87.4%	161	18	0	0	0	0
9	Q9S644	87.4%	76	87.4%	161	87.4%	76	87.4%	161	18	0	0	0	0
10	Q9X3T9	82.8%	72	82.8%	161	82.8%	72	82.8%	161	18	0	0	0	0
11	Q9X3T7	82.8%	72	82.8%	161	82.8%	72	82.8%	161	18	0	0	0	0
12	Q9S643	59.8%	52	59.8%	186	59.8%	52	59.8%	186	18	0	0	0	0
13	P72209	52.9%	46	52.9%	250	52.9%	46	52.9%	250	18	0	0	0	0
14	Q9FXV9	52.9%	46	52.9%	510	52.9%	46	52.9%	510	18	0	0	0	0
15	Q9LRW3	51.7%	45	51.7%	204	51.7%	45	51.7%	204	18	0	0	0	0
16	O87336	51.7%	45	51.7%	216	51.7%	45	51.7%	216	18	0	0	0	0
17	O86218	51.1%	44.5	51.1%	591	51.1%	44.5	51.1%	591	18	0	0	0	0
18	O54204	50.6%	44	50.6%	931	50.6%	44	50.6%	931	18	0	0	0	0

DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE MAJOR PILIN PROTEIN FIMA (FRAGMENT).  
 GN FIMA.  
 OS Salmonella enterica subsp. salamae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxID=59202;  
 RN [1]  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RKS2985;  
 RX MEDLINE=99138752; PubMed=9973358;  
 RA Boyd E.F., Hartl D.L.;  
 RT "Analysis of the type 1 pilin gene cluster firm in Salmonella: its  
 RT distinct evolutionary histories in the 5' and 3' regions."  
 RL J. Bacteriol. 181:1301-1308(1999).  
 DR EMBL; AF083911; AAD23945.1; -;  
 DR InterPro; IPR000259; -;  
 DR Pfam; PF00419; Fimbrin; 1.  
 DR NON\_TER 1  
 FT SEQUENCE 161 AA; 16441 MW; BE3DA852B6144AA2 CRC64;

Query Match 96.6%; Score 84; DB 2; Length 161;  
 Best Local Similarity 94.4%; Pred. No. 1.2e-06;  
 Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFTAGDTTAAQVPSIV 18  
 |||||  
 DB 41 ASFTAGDTTAAQVPSIV 58

RESULT 3

ID O9WM13 PRELIMINARY; PRT; 161 AA.

AC 09WM13;  
 DT 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE MAJOR PILIN PROTEIN FIMA (FRAGMENT).  
 GN FIMA.  
 OS Salmonella enterica subsp. diarizonae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxID=59204;  
 RN [1]  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RKS2979; RKS2978;  
 RX MEDLINE=99138752; PubMed=9973358;  
 RA Boyd E.F., Hartl D.L.;  
 RT "Analysis of the type 1 pilin gene cluster firm in Salmonella: its  
 RT distinct evolutionary histories in the 5' and 3' regions."  
 RL J. Bacteriol. 181:1301-1308(1999).  
 DR EMBL; AF083902; AAD23927.1; -;  
 DR EMBL; AF083901; AAD23925.1; -;  
 DR InterPro; IPR000259; -;  
 DR Pfam; PF00419; Fimbrin; 1.  
 DR NON\_TER 1  
 FT SEQUENCE 161 AA; 16515 MW; 4A7ELEF9ACA5E1B8 CRC64;

Query Match 96.6%; Score 84; DB 2; Length 161;  
 Best Local Similarity 94.4%; Pred. No. 1.2e-06;  
 Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFTAGDTTAAQVPSIV 18  
 |||||  
 DB 41 ASFTAGDTTAAQVPSIV 58

RESULT 4  
 O9X3T5 PRELIMINARY; PRT; 161 AA.

AC 09X3T5;  
 DT 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE MAJOR PILIN PROTEIN FIMA (FRAGMENT).  
 GN FIMA.  
 OS Salmonella enterica subsp. enterica serovar Typhimurium.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxID=90371;  
 RN [1]  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RKS4194;  
 RX MEDLINE=99138752; PubMed=9973358;  
 RA Boyd E.F., Hartl D.L.;  
 RT "Analysis of the type 1 pilin gene cluster firm in Salmonella: its  
 RT distinct evolutionary histories in the 5' and 3' regions."  
 RL J. Bacteriol. 181:1301-1308(1999).  
 DR EMBL; AF083900; AAD23923.1; -;  
 DR InterPro; IPR000259; -;  
 DR Pfam; PF00419; Fimbrin; 1.  
 DR NON\_TER 1  
 FT SEQUENCE 161 AA; 16422 MW; 372FA51E6018939A CRC64;

Query Match 94.3%; Score 82; DB 2; Length 161;  
 Best Local Similarity 94.4%; Pred. No. 2.5e-06;  
 Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFTAGDTTAAQVPSIV 18  
 |||||  
 DB 41 ASFTAGDTTAAQVPSIV 58

RESULT 5  
 ID O53483 PRELIMINARY; PRT; 185 AA.

AC O53483;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)  
 DE FIMA.  
 GN FIMA.  
 OS Salmonella enteritidis.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxID=592;  
 RN [1]  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=95174773; PubMed=7870071;  
 RA Doran J.L., Collinson S.K., Kay C.M., Baner P.A., Burian J.,  
 RA Munro C.K., Lee S.H., Somers J.M., Todd E.C., Kay W.W.;  
 RT "Firm and tctc based DNA diagnostics for Salmonella."  
 RL Mol. Cell. Probes 8:291-310(1994).  
 DR EMBL; S76043; AAB33536.1; -;  
 DR InterPro; IPR000259; -;  
 DR Pfam; PF00419; Fimbrin; 1.  
 DR NON\_TER 1  
 FT SEQUENCE 185 AA; 18870 MW; 1466F94DE872FB87 CRC64;

Query Match 94.3%; Score 82; DB 2; Length 185;  
 Best Local Similarity 94.4%; Pred. No. 2.9e-06;  
 Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ASFTAGDTTAAQVPSIV 18  
 |||||  
 DB 65 ASFTAGDTTAAQVPSIV 82

RESULT 6  
 O9WM63 PRELIMINARY; PRT; 161 AA.

DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE MAJOR PILIN PROTEIN FIWA (FRAGMENT).  
 GN FIMA.  
 OS Salmonella enterica subsp. indica.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 NCBI\_TaxID=59207;  
 OX NCBI\_TaxID=59207;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RKS2955; RKS3057;  
 RX MEDLINE-99138752; PubMed-9973358;  
 RA Boyd E.F., Hartl D.L.;  
 RT "Analysis of the type 1 pilin gene cluster fim in Salmonella: its  
 distinct evolutionary histories in the 5' and 3' regions."  
 RL J. Bacteriol. 181:1301-1308(1999).  
 DR EMBL: AF083912; AAD23947.1; -  
 DR EMBL: AF083905; AAD23933.1; -  
 DR InterPro: IPR000259; -  
 DR Pfam: PF00419; Fimbrin; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 161 AA; 16410 MW; 1E3A02E81D659004 CRC64;

Query Match 90.8%; Score 79; DB 2; Length 161;  
 Best Local Similarity 88.9%; Pred. No. 8.2e-06;  
 Matches 16; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 ASFTAGDTTAQVPSIV 18  
 DB 41 ASFTAGDTTAQVPSIV 58

RESULT 7  
 O9X3U1 PRELIMINARY; PRT; 161 AA.  
 AC O9X3U1;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE MAJOR PILIN PROTEIN FIWA (FRAGMENT).  
 GN FIMA.  
 OS Salmonella enterica subsp. VII.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 NCBI\_TaxID=59208;  
 OX NCBI\_TaxID=59208;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RKS3014;  
 RX MEDLINE-99138752; PubMed-9973358;  
 RA Boyd E.F., Hartl D.L.;  
 RT "Analysis of the type 1 pilin gene cluster fim in Salmonella: its  
 distinct evolutionary histories in the 5' and 3' regions."  
 RL J. Bacteriol. 181:1301-1308(1999).  
 DR EMBL: AF083908; AAD23939.1; -  
 DR InterPro: IPR000259; -  
 DR Pfam: PF00419; Fimbrin; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 161 AA; 16498 MW; C868E19251B6A868 CRC64;

Query Match 87.4%; Score 76; DB 2; Length 161;  
 Best Local Similarity 77.8%; Pred. No. 2.6e-05;  
 Matches 14; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 ASFTAGDTTAQVPSIV 18  
 DB 41 ASFTAGDTTAQVPSIV 58  
 RESULT 8  
 O9WV8

ID O9WV8 PRELIMINARY; PRT; 161 AA.  
 AC O9WV8;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE MAJOR PILIN PROTEIN FIWA (FRAGMENT).  
 GN FIMA.  
 OS Salmonella enterica subsp. houtenae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 NCBI\_TaxID=59205;  
 OX NCBI\_TaxID=59205;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RKS3027; RKS3015;  
 RX MEDLINE-99138752; PubMed-9973358;  
 RA Boyd E.F., Hartl D.L.;  
 RT "Analysis of the type 1 pilin gene cluster fim in Salmonella: its  
 distinct evolutionary histories in the 5' and 3' regions."  
 RL J. Bacteriol. 181:1301-1308(1999).  
 DR EMBL: AF083910; AAD23943.1; -  
 DR EMBL: AF083909; AAD23941.1; -  
 DR InterPro: IPR000259; -  
 DR Pfam: PF00419; Fimbrin; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 161 AA; 16498 MW; C868E19251B6A868 CRC64;

Query Match 87.4%; Score 76; DB 2; Length 161;  
 Best Local Similarity 77.8%; Pred. No. 2.6e-05;  
 Matches 14; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 ASFTAGDTTAQVPSIV 18  
 DB 41 ASFTAGDTTAQVPSIV 58

RESULT 9  
 O9S644 PRELIMINARY; PRT; 161 AA.  
 AC O9S644;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE MAJOR PILIN PROTEIN FIWA (FRAGMENT).  
 GN FIMA.  
 OS Salmonella enterica subsp. VII.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 NCBI\_TaxID=59208;  
 OX NCBI\_TaxID=59208;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RKS3013;  
 RX MEDLINE-99138752; PubMed-9973358;  
 RA Boyd E.F., Hartl D.L.;  
 RT "Analysis of the type 1 pilin gene cluster fim in Salmonella: its  
 distinct evolutionary histories in the 5' and 3' regions."  
 RL J. Bacteriol. 181:1301-1308(1999).  
 DR EMBL: AF083907; AAD23937.1; -  
 DR InterPro: IPR000259; -  
 DR Pfam: PF00419; Fimbrin; 1.  
 DR ProDom: PD000316; -  
 FT NON\_TER 1  
 SQ SEQUENCE 161 AA; 16484 MW; CC39A48251B6A868 CRC64;

Query Match 87.4%; Score 76; DB 2; Length 161;  
 Best Local Similarity 77.8%; Pred. No. 2.6e-05;  
 Matches 14; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 ASFTAGDTTAQVPSIV 18  
 DB 41 ASFTAGDTTAQVPSIV 58

RESULT 10  
ID 09X3T9 PRELIMINARY; PRT: 161 AA.  
AC 09X3T9;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE MAJOR PILIN PROTEIN FIMA (FRAGMENT).  
GN FIMA.  
OS Salmonella enterica subsp. salamae.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=59202;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-RKS2993;  
RX MEDLINE=99138752; PubMed=9973358;  
RA Boyd E.F., Hartl D.L.;  
RT "Analysis of the type 1 pilin gene cluster fim in Salmonella: its  
RT distinct evolutionary histories in the 5' and 3' regions."  
RL J. Bacteriol. 181:1301-1308(1999).  
DR EMBL; AF083906; AAD23935.1; -  
DR InterPro: IPR000259; -  
DR Pfam: PF00419; Fimbrilal; 1.  
FT NON\_TER  
SQ SEQUENCE 161 AA; 16512 MW; C2FF6643E381C7F3 CRC64;

Query Match 82.8%; Score 72; DB 2; Length 161;  
Best Local Similarity 77.8%; Pred. No. 0.00013;  
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
OY 1 ASFTAGDGTAAVPSIV 18  
Db 41 ASFTVGDGTALVPFTIV 58  
|||||:|||||:|||||

RESULT 11  
ID 09X3T7 PRELIMINARY; PRT: 161 AA.  
AC 09X3T7;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE MAJOR PILIN PROTEIN FIMA (FRAGMENT).  
GN FIMA.  
OS Salmonella enterica subsp. arizonae.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=59203;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-RKS2983;  
RX MEDLINE=99138752; PubMed=9973358;  
RA Boyd E.F., Hartl D.L.;  
RT "Analysis of the type 1 pilin gene cluster fim in Salmonella: its  
RT distinct evolutionary histories in the 5' and 3' regions."  
RL J. Bacteriol. 181:1301-1308(1999).  
DR EMBL; AF083904; AAD23951.1; -  
DR InterPro: IPR000259; -  
DR Pfam: PF00419; Fimbrilal; 1.  
FT NON\_TER  
SQ SEQUENCE 161 AA; 16512 MW; C2FF6643E381C7F3 CRC64;

Query Match 82.8%; Score 72; DB 2; Length 161;  
Best Local Similarity 77.8%; Pred. No. 0.00013;  
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
OY 1 ASFTAGDGTAAVPSIV 18  
Db 41 ASFTVGDGTALVPFTIV 58  
|||||:|||||:|||||

Db 41 ASFTVGDGTALVPFTIV 58  
RESULT 12  
ID 09S643 PRELIMINARY; PRT: 161 AA.  
AC 09S643;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE MAJOR PILIN PROTEIN FIMA (FRAGMENT).  
GN FIMA.  
OS Salmonella enterica subsp. arizonae.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=59203;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-RKS2980;  
RX MEDLINE=99138752; PubMed=9973358;  
RA Boyd E.F., Hartl D.L.;  
RT "Analysis of the type 1 pilin gene cluster fim of Salmonella: its  
RT distinct evolutionary histories in the 5' and 3' regions."  
RL J. Bacteriol. 181:1301-1308(1999).  
DR EMBL; AF083903; AAD23929.1; -  
DR InterPro: IPR000259; -  
DR Pfam: PF00419; Fimbrilal; 1.  
FT NON\_TER  
SQ SEQUENCE 161 AA; 16500 MW; C2F9D6455531D456 CRC64;

Query Match 82.8%; Score 72; DB 2; Length 161;  
Best Local Similarity 77.8%; Pred. No. 0.00013;  
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
OY 1 ASFTAGDGTAAVPSIV 18  
Db 41 ASFTVGDGTALVPFTIV 58  
|||||:|||||:|||||

RESULT 13  
ID P72209 PRELIMINARY; PRT: 186 AA.  
AC P72209;  
DT 01-FEB-1997 (TREMBlrel. 02, Created)  
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
DE MAJOR SUBUNIT OF TYPE 1 FIMBRIN.  
GN ATFA.  
OS Proteus mirabilis.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Proteus.  
OX NCBI\_TaxID=584;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H14320;  
RX Massad G., Fulkerson J.F., Watson D.C., Mobley H.L.T.;  
RA Infect. Immun. 64:0-0(1996).  
RL EMBL; Z78535; CAB01712.1; -  
DR InterPro: IPR000259; -  
DR Pfam: PF00419; Fimbrilal; 1.  
SQ SEQUENCE 186 AA; 19061 MW; A2FE208F7D526EC2 CRC64;

Query Match 59.8%; Score 52; DB 2; Length 186;  
Best Local Similarity 58.8%; Pred. No. 0.37;  
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
OY 1 ASFTAGDGTAAVPSI 17  
Db 71 AEFTRKKGDEGRIPFSI 87  
|||||:|||||:|||||

RESULT 14

O9FXY9

AC O9FXY9

PRELIMINARY:

PRT: 250 AA.

DT 01-MAR-2001

(TREMBLrel. 16, Created)

DT 01-MAR-2001

(TREMBLrel. 16, Last sequence update)

DE RESISTANCE

PROTEIN MG55 (FRAGMENT).

OS Glycine max

(Soybean).

OC Eukaryote:

Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta;

eudicotyledons; core eudicots; Rosidae; eurosids I;

OC Fabales;

Fabaceae; Papilionoideae; Glycine.

OX NCBI\_TaxID=3847;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV

WILLIAMS 82; TISSUE=EPICOTYL;

RX MEDLINE=20165944;

PubMed=10701117;

RA Graham M.A., Marek L.F., Lohnes D., Cregan P., Shoemaker R.C.;

RT "Expression and genome organization of resistance gene analogs in

soybean."

RL Genome 43:86-93(2000).

DR EMBL: AF175397; AAC01054.1; -.

FT NON\_TER

FT SEQUENCE

SQ 250 AA; 28320 MW; A4A503ED063A4115 CRC64;

Query Match

Best Local Similarity

Matches 12; Conservative

Score 46; DB 10; Length 250;

Pred. No. 5.4;

Mismatches 0; Indels 3; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Indels 2; Gaps 1;

Db 293 TVISDLTADIPESV 306

Search completed: June 20, 2001, 13:32:42  
Job time: 80 sec

RESULT 15

O9LRW3

ID O9LRW3

AC O9LRW3

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE SIMILARITY TO SCARECROW GENE REGULATOR.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

OC Brassicales; Brassicaceae; Arabidopsids.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;

RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RX MEDLINE=20277480; PubMed=10819329;

RA Nakamura Y.;

RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence

RT features of the regions of 4,504,864 bp covered by sixty pl and TAC

RT clones."

DNA Res. 7:131-135(2000).

EMBL: AB028610; BAB02908.1; -.

SQ SEQUENCE 510 AA; 57181 MW; 559F8C997862866F CRC64;

Query Match

Best Local Similarity

Matches 8; Conservative

Score 46; DB 10; Length 510;

Pred. No. 12;

Mismatches 2; Indels 4; Gaps 0;

Indels 4; Gaps 0;

Indels 4; Gaps 0;

Thu Jun 21 08:38:35 2001

us-09-646-043-1.rpt

Page 6



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 13:39:48 ; Search time 45.83 seconds

(Without alignments)  
13.708 Million cell updates/sec

Title: US-09-646-043-1

Perfect score: 87

Sequence: 1 ASPTAIGDTTAAQVPSIV 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1412450 seqs, 34903136 residues

Total number of hits satisfying chosen parameters: 1412450

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents\_AA\_New:\*\*

1: /cgn2\_6/ptodata/1/paa/PC7\_NEW\_COMB.pcp:\*\*

2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pcp:\*\*

3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pcp:\*\*

4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pcp:\*\*

5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pcp:\*\*

6: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pcp:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87	100.0	18	US-09-646-043-1	Sequence 1, Appl1
2	37	42.5	557	US-09-134-001C-5569	Sequence 5569, Ap
3	36	41.4	224	US-09-134-001C-4010	Sequence 4010, Ap
4	36	41.4	293	US-09-134-001C-3803	Sequence 3803, Ap
5	34	39.1	503	US-09-134-001C-4214	Sequence 4214, Ap
6	33	37.9	333	US-09-454-684A-292	Sequence 292, App
7	33	37.9	333	US-09-620-412A-292	Sequence 292, App
8	33	37.9	476	US-09-134-001C-3778	Sequence 3778, Ap
9	33	37.9	901	US-09-134-001C-5351	Sequence 5351, Ap
10	32	36.8	25	US-09-724-059-562321	Sequence 562321, Ap
11	32	36.8	25	US-09-724-059-564841	Sequence 564841, Ap
12	32	36.8	25	US-09-724-059-567241	Sequence 567241, Ap
13	32	36.8	25	US-09-724-059-569761	Sequence 569761, Ap
14	32	36.8	25	US-09-724-059-574561	Sequence 574561, Ap
15	32	36.8	123	US-09-724-059-579601	Sequence 579601, Ap
16	32	36.8	177	US-09-134-001C-5115	Sequence 5115, Ap
17	32	36.8	248	US-09-134-001C-4482	Sequence 4482, Ap
18	32	36.8	248	US-09-134-001C-5085	Sequence 5085, Ap
19	32	36.8	462	US-09-846-729-3	Sequence 3, Appl1
20	32	36.8	462	US-09-846-729-17	Sequence 17, Appl1
21	32	36.8	464	US-09-846-729-14	Sequence 14, Appl1
22	32	36.8	715	US-09-620-412A-321	Sequence 321, Appl1
23	32	36.8	1693	US-09-851-410-7	Sequence 7, Appl1
24	31	35.6	24	US-09-724-059-445061	Sequence 445061, Ap
25	31	35.6	24	US-09-724-059-447581	Sequence 447581, Ap
26	31	35.6	24	US-09-724-059-449981	Sequence 449981, Ap
27	31	35.6	24	US-09-724-059-452501	Sequence 452501, Ap

28	31	35.6	24	US-09-724-059-457301	Sequence 457301, Ap
29	31	35.6	24	US-09-724-059-462341	Sequence 462341, Ap
30	31	35.6	25	US-09-724-059-572281	Sequence 572281, Ap
31	31	35.6	25	US-09-724-059-577321	Sequence 577321, Ap
32	31	35.6	25	US-09-724-059-582001	Sequence 582001, Ap
33	31	35.6	25	US-09-724-059-584521	Sequence 584521, Ap
34	31	35.6	106	US-09-818-683-364	Sequence 364, App
35	31	35.6	137	US-09-134-001C-3078	Sequence 3078, App
36	31	35.6	218	US-09-497-895-2	Sequence 2, Appl1
37	31	35.6	269	US-09-134-001C-5197	Sequence 5197, App
38	31	35.6	982	US-09-454-684A-176	Sequence 176, App
39	31	35.6	982	US-09-620-412A-176	Sequence 176, App
40	31	35.6	1006	US-09-454-684A-190	Sequence 190, App
41	31	35.6	1006	US-09-620-412A-190	Sequence 190, App
42	30.5	35.1	461	US-09-225-358A-12	Sequence 12, Appl1
43	30.5	35.1	525	US-09-134-001C-3514	Sequence 3514, App
44	30	34.5	24	US-09-724-059-455021	Sequence 455021, Ap
45	30	34.5	24	US-09-724-059-460061	Sequence 460061, Ap

## ALIGNMENTS

RESULT 1

US-09-646-043-1

Sequence 1, Application US/09646043

GENERAL INFORMATION:

APPLICANT: HAKALEHTO, Elino

TITLE OF INVENTION: METHOD FOR DETECTING MICROBES FROM AN ENRICHMENT CULTURE

FILE REFERENCE: 0933-0162P

CURRENT APPLICATION NUMBER: US/09/646, 043

CURRENT FILING DATE: 2000-10-25

NUMBER OF SEQ ID NOS: 1

SOFTWARE: PatentIn version 3.0

SEQ ID NO 1

LENGTH: 18

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Peptide sequence derived from Salmonella typhimurium and E.coli

US-09-646-043-1

Query Match Best Local Similarity 100.0%; Score 87; DB 5; Length 18; Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASPTAIGDTTAAQVPSIV 18  
DB 1 ASPTAIGDTTAAQVPSIV 18

RESULT 2

US-09-134-001C-5569

Sequence 5569, Application US/09134001C

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134, 001C

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055, 779

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 5569

LENGTH: 557

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-5569

Query Match 42.5%; Score 37; DB 5; Length 557;  
Best Local Similarity 50.0%; Pred. No. 8.8e+02;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 3 FTAGDTPAQPVSIV 14  
| : | : | : | : | :  
DB 190 FEALGDTSEMP 201

RESULT 3  
US-09-134-001C-4010  
; Sequence 4010, Application US/09134001C  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4010  
; LENGTH: 224  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4010

Query Match 41.4%; Score 36; DB 5; Length 224;  
Best Local Similarity 50.0%; Pred. No. 4.7e+02;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 4 TAIGDTPAQPVSIV 17  
| : | : | : | : | :  
DB 92 TSIGTAAIVPLTV 105

RESULT 4  
US-09-134-001C-3803  
; Sequence 3803, Application US/09134001C  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 3803  
; LENGTH: 293  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3803

Query Match 41.4%; Score 36; DB 5; Length 293;  
Best Local Similarity 33.3%; Pred. No. 6.4e+02;  
Matches 6; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 1 ASFTAGDTPAQPVSIV 18  
| : | : | : | : | :  
DB 219 ASFASVGDINLSFKALI 236

RESULT 5  
US-09-134-001C-4214  
; Sequence 4214, Application US/09134001C  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4214  
; LENGTH: 503  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4214

Query Match 39.1%; Score 34; DB 5; Length 503;  
Best Local Similarity 40.0%; Pred. No. 2.7e+03;  
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 4 TAIGDTPAQPVSIV 18  
| : | : | : | : | :  
DB 321 TQAGDISATVPNTVI 335

RESULT 6  
US-09-454-684A-292  
; Sequence 292, Application US/09454684A  
; GENERAL INFORMATION:  
; APPLICANT: Probst, Peter  
; APPLICANT: Bhatia, Ajay  
; APPLICANT: Skeiky, Yasir  
; APPLICANT: Fling, Steve  
; APPLICANT: Maisonneuve, Jeff  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; FILE REFERENCE: 210121.469C4  
; CURRENT APPLICATION NUMBER: US/09/454,684A  
; CURRENT FILING DATE: 1999-12-03  
; NUMBER OF SEQ ID NOS: 310  
; SOFTWARE: PASTSRQ for Windows Version 3.0/4.0  
; SEQ ID NO 292  
; LENGTH: 333  
; TYPE: PRT  
; ORGANISM: Chlamydia  
US-09-454-684A-292

Query Match 37.9%; Score 33; DB 5; Length 333;  
Best Local Similarity 40.0%; Pred. No. 2.5e+03;  
Matches 6; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 4 TAIGDTPAQPVSIV 18  
| : | : | : | : | :  
DB 167 TAVGATTSATFGVL 181

RESULT 7  
US-09-620-412A-292  
; Sequence 292, Application US/09620412A  
; GENERAL INFORMATION:  
; APPLICANT: Steven P. Fling  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; FILE REFERENCE: 210121.469C7  
; CURRENT APPLICATION NUMBER: US/09/620,412A  
; CURRENT FILING DATE: 2001-07-20

NUMBER OF SEQ ID NOS: 357  
SOFTWARE: FastSeq for Windows Version 3.0/4.0  
SEQ ID NO 292  
LENGTH: 333  
TYPE: PRT  
ORGANISM: Chlamydia  
US-09-620-412A-292

Query Match  
Best Local Similarity 37.9%; Score 33; DB 5; Length 333;  
Sequence 3778, Application US/09134001C  
Matches 6; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 4 FTAGDTAQPFSIV 18  
DB 167 FAVGATYTSATFGVL 161

RESULT 8  
US-09-134-001C-3778  
Sequence 3778, Application US/09134001C  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 3778  
LENGTH: 476  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3778

Query Match  
Best Local Similarity 37.9%; Score 33; DB 5; Length 476;  
Sequence 3778, Application US/09134001C  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 7 GDTTAQVPSFIV 18  
DB 294 GDSIATSPFTLV 305

RESULT 9  
US-09-134-001C-5351  
Sequence 5351, Application US/09134001C  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 5351  
LENGTH: 901  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5351

Query Match  
Best Local Similarity 37.9%; Score 33; DB 5; Length 901;  
Sequence 5351, Application US/09134001C  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
OY 2 SFATGDT 9  
DB 467 NFTAIGDS 474

RESULT 10  
US-09-724-059-562321  
Sequence 562321, Application US/09724059  
GENERAL INFORMATION:  
APPLICANT: Choo, Yen  
APPLICANT: Kling, Aaron  
TITLE OF INVENTION: Nucleic Acid Binding Proteins  
FILE REFERENCE: P2500USM  
CURRENT APPLICATION NUMBER: US/09/724,059  
CURRENT FILING DATE: 2000-11-28  
PRIOR APPLICATION NUMBER: GB9710809.6  
PRIOR FILING DATE: 1997-05-23  
NUMBER OF SEQ ID NOS: 1407122  
SOFTWARE: MacAlister  
SEQ ID NO 562321  
LENGTH: 25  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Polypeptide sequence  
US-09-724-059-562321

Query Match  
Best Local Similarity 36.8%; Score 32; DB 5; Length 25;  
Sequence 562321, Application US/09724059  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 FTAIGDTTA 11  
DB 12 FTRLGDLTA 20

RESULT 11  
US-09-724-059-564841  
Sequence 564841, Application US/09724059  
GENERAL INFORMATION:  
APPLICANT: Choo, Yen  
APPLICANT: Kling, Aaron  
TITLE OF INVENTION: Nucleic Acid Binding Proteins  
FILE REFERENCE: P2500USM  
CURRENT APPLICATION NUMBER: US/09/724,059  
CURRENT FILING DATE: 2000-11-28  
PRIOR APPLICATION NUMBER: GB9710809.6  
PRIOR FILING DATE: 1997-05-23  
NUMBER OF SEQ ID NOS: 1407122  
SOFTWARE: MacAlister  
SEQ ID NO 564841  
LENGTH: 25  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Polypeptide sequence  
US-09-724-059-564841

Query Match  
Best Local Similarity 36.8%; Score 32; DB 5; Length 25;  
Sequence 564841, Application US/09724059  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 FTAIGDTTA 11  
DB 12 FTRLGDLTA 20

RESULT 12  
US-09-724-059-567241  
; Sequence 567241, Application US/09724059  
; GENERAL INFORMATION:  
; APPLICANT: Choo, Yen  
; APPLICANT: Klug, Aaron  
; APPLICANT: Isalan, Mark  
; TITLE OF INVENTION: Nucleic Acid Binding Proteins  
; FILE REFERENCE: P2500USM  
; CURRENT APPLICATION NUMBER: US/09/724,059  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: GB9710809.6  
; PRIOR FILING DATE: 1997-05-23  
; NUMBER OF SEQ ID NOS: 1407122  
; SOFTWARE: Macallister  
; SEQ ID NO 567241  
; LENGTH: 25  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Polypeptide sequence  
US-09-724-059-567241

Query Match 36.8%; Score 32; DB 5; Length 25;  
Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 FTAIGDTTA 11  
11:11111  
DB 12 FTQLGDLTA 20

RESULT 13  
US-09-724-059-569761  
; Sequence 569761, Application US/09724059  
; GENERAL INFORMATION:  
; APPLICANT: Choo, Yen  
; APPLICANT: Klug, Aaron  
; APPLICANT: Isalan, Mark  
; TITLE OF INVENTION: Nucleic Acid Binding Proteins  
; FILE REFERENCE: P2500USM  
; CURRENT APPLICATION NUMBER: US/09/724,059  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: GB9710809.6  
; PRIOR FILING DATE: 1997-05-23  
; NUMBER OF SEQ ID NOS: 1407122  
; SOFTWARE: Macallister  
; SEQ ID NO 569761  
; LENGTH: 25  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Polypeptide sequence  
US-09-724-059-569761

Query Match 36.8%; Score 32; DB 5; Length 25;  
Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 FTAIGDTTA 11  
11:11111  
DB 12 FTQLGDLTA 20

RESULT 14  
US-09-724-059-574561  
; Sequence 574561, Application US/09724059  
; GENERAL INFORMATION:  
; APPLICANT: Choo, Yen  
; APPLICANT: Klug, Aaron  
; APPLICANT: Isalan, Mark

; TITLE OF INVENTION: Nucleic Acid Binding Proteins  
; FILE REFERENCE: P2500USM  
; CURRENT APPLICATION NUMBER: US/09/724,059  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: GB9710809.6  
; PRIOR FILING DATE: 1997-05-23  
; NUMBER OF SEQ ID NOS: 1407122  
; SOFTWARE: Macallister  
; SEQ ID NO 574561  
; LENGTH: 25  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Polypeptide sequence  
US-09-724-059-574561

Query Match 36.8%; Score 32; DB 5; Length 25;  
Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 FTAIGDTTA 11  
11:11111  
DB 12 FTQLGDLTA 20

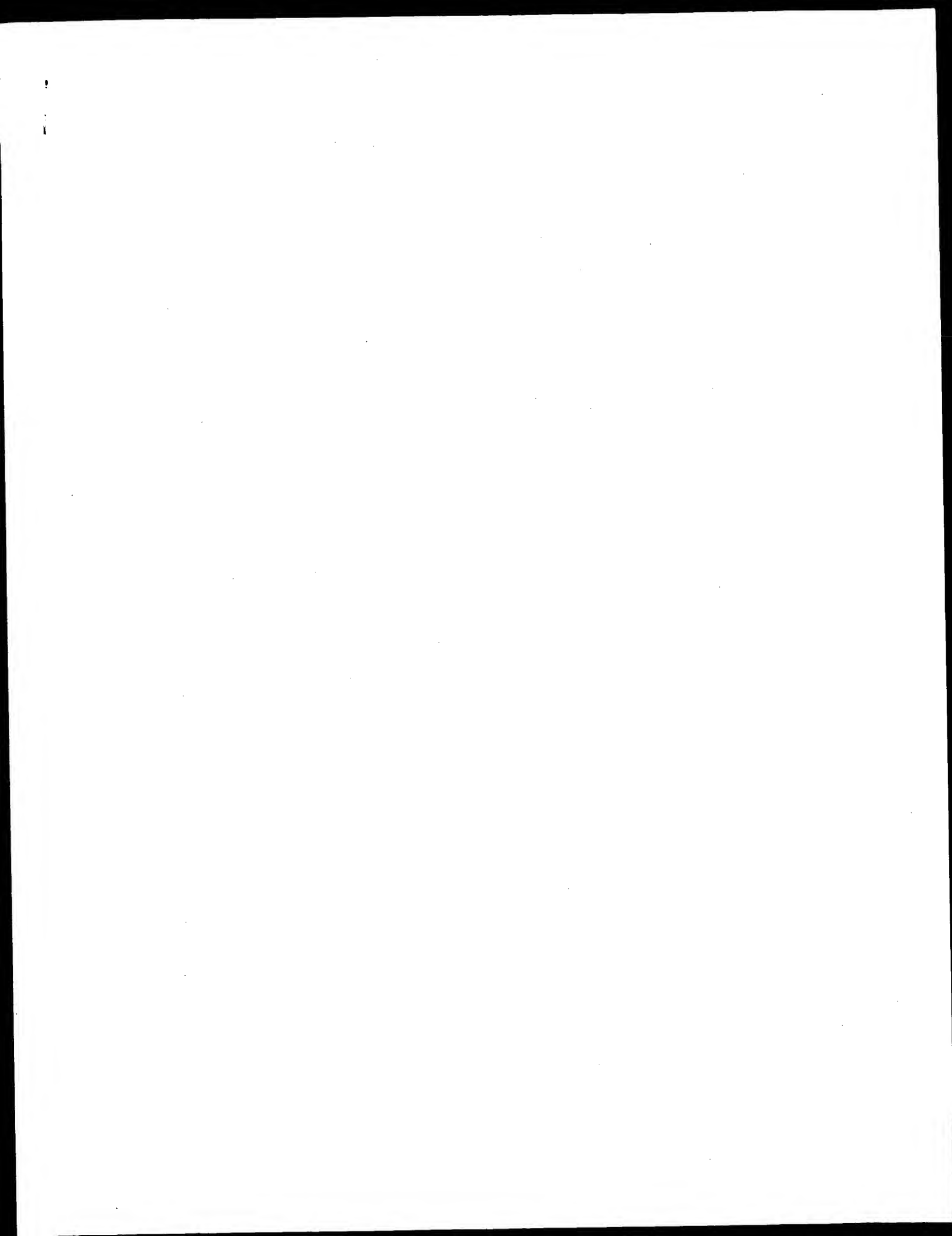
RESULT 15  
US-09-724-059-579601  
; Sequence 579601, Application US/09724059  
; GENERAL INFORMATION:  
; APPLICANT: Choo, Yen  
; APPLICANT: Klug, Aaron  
; APPLICANT: Isalan, Mark  
; TITLE OF INVENTION: Nucleic Acid Binding Proteins  
; FILE REFERENCE: P2500USM  
; CURRENT APPLICATION NUMBER: US/09/724,059  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: GB9710809.6  
; PRIOR FILING DATE: 1997-05-23  
; NUMBER OF SEQ ID NOS: 1407122  
; SOFTWARE: Macallister  
; SEQ ID NO 579601  
; LENGTH: 25  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Polypeptide sequence  
US-09-724-059-579601

Query Match 36.8%; Score 32; DB 5; Length 25;  
Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 FTAIGDTTA 11  
11:11111  
DB 12 FTQLGDLTA 20

Search completed: June 20, 2001, 13:42:32  
Job time: 164 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 13:38:53 ; Search time 44.73 Seconds

(without alignments)  
71.197 Million cell updates/sec

Title: US-09-646-043-1

Sequence: 1 ASFTAGDTRTAQVPSIV 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1316263 seqs, 17692484 residues

Total number of hits satisfying chosen parameters: 1316263

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_Main:

1: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
7: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
8: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
9: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
10: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
11: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
12: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
13: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
14: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
15: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
16: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
17: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
18: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
19: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
20: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
21: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
22: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*  
23: /cgn2\_6/ptodata/2/paa/US086.COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	61	70.1	137	16	US-09-252-691-6209
2	61	70.1	137	16	US-09-252-691-6209
3	47	54.0	171	23	US-09-711-164-359
4	47	54.0	171	23	US-09-711-164-359
5	47	54.0	171	23	US-09-711-164-359
6	47	54.0	171	23	US-09-711-164-359
7	46	52.9	348	21	US-09-489-039A-13057
8	43	49.4	437	16	US-09-733-089-16780
9	43	49.4	437	16	US-09-733-089-16780
10	42	48.3	205	6	US-08-277-231-5

11	42	48.3	442	19	US-09-540-236-1728	Sequence 3728, Ap
12	42	48.3	442	23	US-60-128-476-4764	Sequence 4764, Ap
13	41	47.1	417	18	US-09-489-039A-11753	Sequence 11753, A
14	41	47.1	504	1	PCT-US00-33549-47	Sequence 47, Appl
15	41	47.1	1101	16	US-09-248-796-17347	Sequence 17347, A
16	40	46.0	152	20	US-09-615-846-13	Sequence 13, Appl
17	40	46.0	388	23	US-60-140-956-1369	Sequence 1369, Ap
18	40	46.0	388	23	US-60-143-992-971	Sequence 971, Ap
19	40	46.0	1732	14	US-09-066-330-11	Sequence 11, Appl
20	40	46.0	1732	15	US-09-174-517-10	Sequence 14, Appl
21	40	46.0	2628	15	US-09-174-517-10	Sequence 40941, A
22	39	44.8	51	16	US-09-270-767-40941	Sequence 56157, A
23	39	44.8	51	16	US-09-270-767-40941	Sequence 194448, A
24	39	44.8	134	1	PCT-US01-01435-108	Sequence 108, App
25	39	44.8	164	21	US-09-733-089-3386	Sequence 3386, Ap
26	39	44.8	201	1	PCT-US01-01435-108	Sequence 77, Appl
27	39	44.8	201	1	PCT-US01-01435-108	Sequence 1606, Ap
28	39	44.8	201	1	PCT-US01-01435-108	Sequence 1606, Ap
29	39	44.8	201	1	PCT-US01-01435-108	Sequence 20, Appl
30	39	44.8	238	1	PCT-US00-32990-20	Sequence 20, Appl
31	39	44.8	238	1	PCT-US00-32990-20	Sequence 20, Appl
32	39	44.8	266	18	US-09-417-507-37388	Sequence 37388, A
33	39	44.8	280	23	US-60-230-435-1062	Sequence 1062, Ap
34	39	44.8	332	1	PCT-US00-07535-119	Sequence 119, App
35	39	44.8	346	18	US-09-417-507-41578	Sequence 41578, A
36	39	44.8	393	1	PCT-US00-06049-125	Sequence 125, App
37	39	44.8	393	1	PCT-US00-14973-113	Sequence 113, App
38	39	44.8	393	1	PCT-US00-15137-113	Sequence 113, App
39	39	44.8	393	1	PCT-US00-26324-100	Sequence 100, App
40	39	44.8	404	20	US-09-603-1248-214	Sequence 214, App
41	39	44.8	411	23	US-60-167-217-22819	Sequence 22819, A
42	39	44.8	411	23	US-60-191-637-22589	Sequence 22589, A
43	39	44.8	434	23	US-60-191-637-22589	Sequence 22589, A
44	39	44.8	434	23	US-60-191-637-22589	Sequence 17843, A
45	39	44.8	634	23	US-60-212-659-550	Sequence 550, App

## ALIGNMENTS

RESULT 1

US-09-252-691-6209

Sequence 6209, Application US/09252691B

GENERAL INFORMATION:

APPLICANT: Keith G. Weinstein et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTE

FILE REFERENCE: 107196.135

CURRENT APPLICATION NUMBER: US/09/252,691B

CURRENT FILING DATE: 1999-02-18

NUMBER OF SEQ ID NOS: 11324

SEQ ID NO 6209

LENGTH: 137

TYPE: PRT

ORGANISM: Enterobacter cloacae

US-09-252-691-6209

Query Match 70.1%; Score 61; DB 16; Length 137;

Best Local Similarity 58.8%; Pred. No. 0.0054;

Matches 10; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

DB 17 AKFTKVDITSNIPFI 33

US-09-252-691C-6209

Sequence 6209, Application US/09252691C

GENERAL INFORMATION:

APPLICANT: Keith G. Weinstein et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTE

```

; TITLE OF INVENTION: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.135
; CURRENT FILING DATE: 1999-02-18
; CURRENT APPLICATION NUMBER: US 60/094,145
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 60/074,787
; PRIOR FILING DATE: 1998-02-18
; NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 6209
; LENGTH: 137
; TYPE: PRT
; ORGANISM: Enterobacter cloacae
US-09-252-691C-6209

Query Match          70.1%; Score 61; DB 16; Length 137;
Best Local Similarity 58.8%; Pred. No. 0.0054;
Matches 10: Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 ASFTAGDTTAYPFESI 17
   |||:||||:|-|-|-|
Db 17 AKFTKVGDTSNIPFTI 33

RESULT 3
US-09-711-164-359
; Sequence 359, Application US/09711164
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlson, Kari
; APPLICANT: Judson, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY
; FILE REFERENCE: ELITRA-008A
; CURRENT FILING DATE: 2000-11-09
; CURRENT APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 359
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-711-164-359

Query Match          54.0%; Score 47; DB 21; Length 171;
Best Local Similarity 81.8%; Pred. No. 2.5;
Matches 9: Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 7 GDTTAYPFESI 17
   |||||:|||||
Db 65 GDTTALVPFSL 75

RESULT 4
US-60-164-415-359
; Sequence 359, Application US/60164415
; GENERAL INFORMATION:
; APPLICANT: Bauer, Paula
; APPLICANT: Forsyth, Allyn
; APPLICANT: Froelich, Jamie
; APPLICANT: McCarthy, Melissa
; APPLICANT: Ohlson, Kari
; APPLICANT: Phuong, Trung
; APPLICANT: Robbins, David
; APPLICANT: Tan, Zehui
; APPLICANT: Trawick, John
; APPLICANT: Turcek, Bobbi
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY
; FILE REFERENCE: ELITRA-008PR
```

```

; CURRENT APPLICATION NUMBER: US/60/164,415
; CURRENT FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 359
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Escherichia coli
US-60-164-415-359

Query Match          54.0%; Score 47; DB 23; Length 171;
Best Local Similarity 81.8%; Pred. No. 2.5;
Matches 9: Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 7 GDTTAYPFESI 17
   |||||:|||||
Db 65 GDTTALVPFSL 75

RESULT 5
US-09-489-039A-13057
; Sequence 13057, Application US/09489039A
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT FILING DATE: 2000-01-27
; CURRENT APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13057
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13057

Query Match          54.0%; Score 47; DB 18; Length 173;
Best Local Similarity 52.9%; Pred. No. 2.5;
Matches 9: Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 ASFTAGDTTAYPFESI 17
   |||:||||:|-|-|-|
Db 59 AOEFTVGTTLVPFPAI 75

RESULT 6
US-09-489-039A-12007
; Sequence 12007, Application US/09489039A
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT FILING DATE: 2000-01-27
; CURRENT APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12007
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12007

Query Match          54.0%; Score 47; DB 18; Length 212;
Best Local Similarity 62.5%; Pred. No. 3.2;
Matches 10: Conservative 1; Mismatches 5; Indels 0; Gaps 0;
```



OY 3 FTAGDTTAQVPSIV 18  
| : | | | | |  
Db 95 FVAVGATTPVPRIV 110

RESULT 7  
US-09-733-089-16780  
Sequence 16780, Application US/09733089  
GENERAL INFORMATION:  
APPLICANT: Dotson, Stanton B.  
APPLICANT: Kovalic, David K.  
APPLICANT: Liu, Jingdong  
APPLICANT: Lutfiyar, Linda L.  
APPLICANT: McIninch, James  
APPLICANT: Wu, Wei  
TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
FILE REFERENCE: 38-21(1500)D  
CURRENT APPLICATION NUMBER: US/09/733,089  
PRIOR FILING DATE: 2000-12-11  
PRIOR APPLICATION NUMBER: US 09/474,435  
PRIOR FILING DATE: 1999-12-28  
PRIOR APPLICATION NUMBER: US 09/654,617  
PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: US 09/620,392  
NUMBER OF SEQ ID NOS: 24143  
SEQ ID NO 16780  
LENGTH: 348  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
US-09-733-089-16780

Query Match 52.9%; Score 46; DB 21; Length 348;  
Best Local Similarity 57.1%; Pred. No. 8.7;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
OY 4 TAIGDTTAQVPSI 17  
| : | | | | |  
Db 293 TVISDLTADIPFSV 306

RESULT 8  
US-09-248-796-19640  
Sequence 19640, Application US/09248796  
GENERAL INFORMATION:  
APPLICANT: Ketch Weinstein et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
FILE REFERENCE: 107196.132  
CURRENT APPLICATION NUMBER: US/09/248,796  
CURRENT FILING DATE: 1999-02-12  
NUMBER OF SEQ ID NOS: 28206  
SEQ ID NO 19640  
LENGTH: 437  
TYPE: PRT  
ORGANISM: Candida albicans  
US-09-248-796-19640

Query Match 49.4%; Score 43; DB 16; Length 437;  
Best Local Similarity 42.9%; Pred. No. 40;  
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;  
OY 4 TAIGDTTAQVPSI 17  
| : | | | | |  
Db 128 SSVGDTTQLPFQV 141

RESULT 9  
US-60-238-332-2  
Sequence 2, Application US/60238332

GENERAL INFORMATION:  
APPLICANT: Lee, Ernestine A.  
APPLICANT: Rankumar, Jaya  
APPLICANT: Yao, Monique G.  
APPLICANT: Patterson, Chandra  
APPLICANT: Walla, Narinder K.  
APPLICANT: Lu, Yan  
APPLICANT: Tribouley, Catherine M.  
APPLICANT: Yue, Henry  
APPLICANT: Wang, Yumel E.  
APPLICANT: Gandhi, Ameena R.  
APPLICANT: Hafalla, April  
APPLICANT: Xu, Yuming  
APPLICANT: Thangavelu, Kavitha  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: PROTEIN PHOSPHATASES  
FILE REFERENCE: PI-0248 P  
CURRENT APPLICATION NUMBER: US/60/238,332  
CURRENT FILING DATE: 2000-10-05  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PERL Program  
SEQ ID NO 2  
LENGTH: 986  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
OTHER INFORMATION: Incycle ID No: 511906CD1  
US-60-238-332-2

Query Match 49.4%; Score 43; DB 23; Length 986;  
Best Local Similarity 50.0%; Pred. No. 1e+02;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
OY 2 SFTAGDTTAQVPSI 17  
| : | | | | |  
Db 74 TETAGADSPQIPFSL 89

RESULT 10  
US-08-277-231-5  
Sequence 5, Application US/08277231  
GENERAL INFORMATION:  
APPLICANT: Green, Bruce A.  
APPLICANT: Brinton, Charles C.  
TITLE OF INVENTION: Sequence and Analysis of LKP P111n  
TITLE OF INVENTION: Structural Genes and The LKP P111 Operon of Nontypable  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millitia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/277,231  
FILING DATE: 19-JUL-1994  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Brook, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: ACC94-02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540

## RESULT 13

GENERAL NAME:  
APPLICANT: Keith Weinstock et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALB

Thu Jun 21 08:38:32 2001

us-09-646-043-1.rapm

Page 5

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796  
; CURRENT FILING DATE: 1999-02-12  
; NUMBER OF SEQ ID NOS: 28206  
; SEQ ID NO 17347  
; LENGTH: 1101  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796-17347

Query Match 47.18; Score 41; DB 16; Length 1101;  
Best Local Similarity 60.08; Pred. No. 2.7e+02;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
OY 1 ASPTAIGDTTAAQVVF 15  
:||||:| ||| |  
DB 993 SSTAVTDFRQVDF 1007

Search completed: June 20, 2001, 13:41:43  
Job time: 170 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 20, 2001, 13:31:22 ; Search time 18.33 Seconds

(without alignments)  
56.134 Million cell updates/sec

Title: US-09-646-043-1

Perfect score: 87

Sequence: 1 ASPTAIGDTTACVPSIV 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_0401.\*

1: /SID56/gcgdata/geneseq/geneseqp/AA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseqp/AA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseqp/AA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseqp/AA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseqp/AA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseqp/AA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseqp/AA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseqp/AA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseqp/AA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseqp/AA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseqp/AA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseqp/AA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseqp/AA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseqp/AA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseqp/AA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseqp/AA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseqp/AA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseqp/AA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseqp/AA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseqp/AA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseqp/AA2000.DAT.\*  
22: /SID56/gcgdata/geneseq/geneseqp/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87	100.0	18	V29989	Salmonella type 1
2	82	94.3	185	W23564	Salmonella enterit
3	42	48.3	205	R89328	Minor tip-associat
4	42	48.3	1120	Y19780	B. burgdorferi ant
5	42	48.3	1146	Y19779	B. burgdorferi ant
6	41	47.1	215	B16004	E. coli proliferat
7	40	46.0	176	B16007	E. coli proliferat
8	40	46.0	1732	R36029	P. gingivalis porp
9	40	46.0	1732	W24787	Prk antigenic pro
10	40	46.0	1732	W69487	Haemagglutinin pro
11	40	46.0	2628	R96030	P. gingivalis haem

12	40	46.0	2628	19	W69488
13	39	44.8	248	21	B41673
14	39	44.8	332	21	B34831
15	39	44.8	393	21	B27624
16	39	44.8	393	22	B60757
17	39	44.8	516	18	W17974
18	38	43.7	303	21	R84825
19	38	43.7	327	21	G45702
20	38	43.7	467	21	G45701
21	38	43.7	467	21	G45700
22	38	43.7	646	18	W13470
23	38	43.7	1135	15	R50037
24	37	42.5	46	21	G25978
25	37	42.5	117	21	Y74376
26	37	42.5	118	22	B63602
27	37	42.5	226	21	B53207
28	37	42.5	259	21	B26244
29	37	42.5	1876	16	R72853
30	37	42.5	1876	16	R72853
31	37	42.5	1895	16	R72854
32	36	41.4	106	20	Y25639
33	36	41.4	132	20	Y25641
34	36	41.4	138	20	Y25636
35	36	41.4	138	20	Y25640
36	36	41.4	138	20	Y25640
37	36	41.4	207	21	B44594
38	36	41.4	255	20	Y34658
39	36	41.4	278	21	G06910
40	36	41.4	292	21	G40208
41	36	41.4	535	21	G29571
42	36	41.4	546	21	G29570
43	36	41.4	681	17	R92701
44	36	41.4	702	21	G29569
45	36	41.4	830	16	B53128
					R65216

## ALIGNMENTS

RESULT	ID	Y29989	standard; peptide: 18 AA.
XX	XX	Y29989;	
AC	XX	24-NOV-1999	(first entry)
DT	XX	Salmonella type 1	fimbriae peptide.
XX	XX	Salmonella; type 1	fimbriae; immunisation; detection; microbe;
DE	XX	enteric bacterium;	microbiological.
XX	XX	Synthetic.	
OS	XX	Salmonella typhimurium.	
OS	XX	Escherichia coli.	
XX	XX	W09947931-AA.	
PN	XX	23-SEP-1999.	
PD	XX	15-MAR-1999;	99WO-F100192.
XX	XX	13-MAR-1998;	98FT-0000571.
PA	XX	(HAKA/)	HAKALEHTO E. E.
XX	XX	Hakalehto EE;	
PI	XX	WPI; 1999-562222/47.	
DR	XX	Microbiological determination, useful for detecting microbes in	
XX	XX	clinical samples, food and environmental samples	
PT	XX		
XX	XX		

Applied Biosystems

aa	Sequence	185 AA
SQ		

1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

```
Qy      1 ASFTAIQDTAOPFSI 17
        | : | | | | |
Db      96 anishagqtaapvpsl 112
```



XX DR WPI: 2000-514822/46.  
XX DR N-PSDB: A66010.  
XX PT Novel polynucleotides and polypeptides associated with microorganism  
XX proliferation, used to identify inhibitors of bacterial growth and  
XX proliferation, for use in antisense therapy -  
XX  
XX Claim 11; Page 270; 316pp; English.  
XX  
XX A65809 to A65889 and A66058 to A66138 represent nucleotide sequences  
XX derived from *Escherichia coli* which inhibit *E. coli* proliferation.  
XX A65890 to A66055 and B15886 to B16040 represent nucleotide and protein  
XX sequences associated with *E. coli* proliferation. A66056 and A66057  
XX represent primers used for sequencing *E. coli* proliferation inhibiting  
XX nucleotide inserts in an example from the present invention. Methods  
XX from the present invention can be used to identify a proliferation-  
XX required gene in a microorganism, by contacting a microorganism with a  
XX proliferation-required gene activity inhibitory nucleic acid identified  
XX in another organism, and determining if inhibition occurs in the second  
XX microorganism. The nucleic acid sequences identified as being required  
XX for bacterial growth and proliferation, can be used for antisense  
XX therapy for killing bacteria.  
XX  
XX Sequence 215 AA:  
SQ  
Query Match 47.1%; Score 41; DB 21; Length 215;  
Best Local Similarity 53.3%; Pred. No. 11;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
OY 3 FTALGDTTAQVPFSI 17  
| : : : : |  
Db 100 fhavgdsapvptvi 114  
RESULT 7  
ID B16007 standard; Protein: 176 AA.  
XX  
XX B16007;  
XX  
XX 05-OCT-2000 (first entry)  
XX  
XX E. coli proliferation associated protein sequence SEQ ID NO:365.  
XX  
XX *Escherichia coli*; *E. coli*; proliferation; inhibition; screening;  
XX antimicrobial; bacterial growth; antisense therapy; antibacterial.  
XX  
XX *Escherichia coli*.  
XX  
XX WO200044906-A2.  
XX  
XX 03-AUG-2000.  
XX  
XX 27-JAN-2000; 2000WO-US02200.  
XX  
XX 27-JAN-1999; 9905-0117405.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr CJ,  
XX Yamamoto RT, Xu HH;  
XX WPI: 2000-514822/46.  
XX N-PSDB: A66013.  
XX  
XX Novel polynucleotides and polypeptides associated with microorganism  
XX proliferation, used to identify inhibitors of bacterial growth and  
XX proliferation, for use in antisense therapy -  
XX  
XX Claim 11; Page 273-274; 316pp; English.  
XX

CC A65809 to A65889 and A66058 to A66138 represent nucleotide sequences  
CC derived from *Escherichia coli* which inhibit *E. coli* proliferation.  
CC A65890 to A66055 and B15886 to B16040 represent nucleotide and protein  
CC sequences associated with *E. coli* proliferation. A66056 and A66057  
CC represent primers used for sequencing *E. coli* proliferation inhibiting  
CC nucleotide inserts in an example from the present invention. Methods  
CC from the present invention can be used to identify a proliferation-  
CC required gene in a microorganism, by contacting a microorganism with a  
CC proliferation-required gene activity inhibitory nucleic acid identified  
CC in another organism, and determining if inhibition occurs in the second  
CC microorganism. The nucleic acid sequences identified as being required  
CC for bacterial growth and proliferation, can be used for antisense  
CC therapy for killing bacteria.  
CC  
CC Sequence 176 AA;  
SQ  
Query Match 46.0%; Score 40; DB 21; Length 176;  
Best Local Similarity 56.2%; Pred. No. 13;  
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
OY 3 FTALGDTTAQVPFSIV 18  
| : : : : |  
Db 59 fmgattpvptvptll 74  
RESULT 8  
ID R96029 standard; Protein: 1732 AA.  
XX  
XX R96029;  
XX  
XX 04-SEP-1996 (first entry)  
XX  
XX P. gingivalis porphyrin.  
XX  
XX Porphyrin; haemagglutinin; periodontal disease; vaccine; antibody.  
XX  
XX Porphyromonas gingivalis strain W12.  
XX  
XX  
XX OS  
XX  
XX Key Location/Qualifiers  
XX  
XX FT Region 688..708 repeat region type 1"  
XX FT /note="Pro-Asn repeat region type 1"  
XX FT 887..952  
XX FT /note="Pro-Asn repeat region type 2"  
XX FT 946..967  
XX FT /note="Pro-Asn repeat region type 1"  
XX FT 985..1006  
XX FT /note="Pro-Asn repeat region type 3"  
XX FT 1041..1100  
XX FT /note="Pro-Asn repeat region type 4"  
XX FT 1341..1405  
XX FT /note="Pro-Asn repeat region type 2"  
XX FT 1430..1451  
XX FT /note="Pro-Asn repeat region type 3"  
XX FT 1488..1547  
XX FT /note="Pro-Asn repeat region type 4"  
XX FT 1607..1650  
XX FT /note="Pro-Asn repeat region type 2"  
XX  
XX WO9617936-A2.  
XX  
XX 13-JUN-1996.  
XX  
XX 11-DEC-1995; 95WO-US16108.  
XX  
XX 09-DEC-1994; 94US-0353485.  
XX  
XX (UABR-) UAB RES FOUND.  
XX (UYFL) UNIV FLORIDA.  
XX  
XX Han N, Lantz M, Lepine G, Patil JM, Progniske-Fox A,  
XX Tumwasorn S;  
PI



XX DR WPI: 1996-287181/29.  
 XX DR N-PSDB; T30653.  
 XX PT Porphyromonas gingivalis genes and proteins - used in the detection  
 XX PT and vaccination against periodontal disease  
 XX PS Claim 5; Page 76-81; 153pp; English.  
 CC CC P. gingivalis W12 cysteine protease, porphyrin (R96029), was  
 CC identified as the product of the prt gene (T30653) isolated from  
 CC P. gingivalis W12 genomic DNA. The porphyrin shows homology to  
 CC the haemagglutinin (see also R96026-28 and R96030-33) of P.  
 CC gingivalis 318. It can be obtained from transformed host cells and  
 CC used as a vaccine to protect humans or animals against periodontal  
 CC disease. Expression in *Salmonella* cells allows production of a live  
 CC vaccine. The porphyrin and haemagglutinins can also be used to  
 CC detect the presence of anti-P. gingivalis antibodies and to raise  
 CC monoclonal antibodies for diagnostic application.  
 XX SQ Sequence 1732 AA;  
 OY 2 SFPAIGDTTAQVPEFI 17  
 Db 1658 ntsladvtaqkpytl 1673  
 RESULT 9  
 W24787  
 ID W24787 standard; Protein; 1732 AA.  
 AC W24787;  
 XX 25-NOV-1997 (first entry)  
 DE PrtK antigenic protein complex.  
 XX Periodontal disease: cell surface protein; thiol protease;  
 KW endopeptidase; PrtK; PrtK48; PrtK35; PrtK15; PrtK44;  
 KW haemagglutinin; adhesin; therapy; diagnosis; vaccine; antigen.  
 XX OS Porphyromonas gingivalis strain W50.  
 XX Key  
 FH Location/Qualifiers  
 FT Peptide  
 FT 1..228  
 FT /label= Pro-pro-peptide  
 FT 228..229  
 FT Protein  
 FT /label= PrtK48  
 FT /note= "48 kDa Lys-specific thiol protease"  
 FT 737..738  
 FT Cleavage-site  
 FT 738..1156  
 FT /label= PrtK39  
 FT /note= "39 kDa adhesin"  
 FT 1156..1157  
 FT Protein  
 FT /label= PrtK15  
 FT /note= "15 kDa adhesin"  
 FT 1291..1292  
 FT Cleavage-site  
 FT 1292..1732  
 FT Protein  
 FT /label= PrtK44  
 FT /note= "44 kDa adhesin"  
 XX PN W09716542-A1.  
 XX PD 09-MAY-1997.  
 XX PF 30-OCT-1996; 96WO-AU00673.

XX XX 30-OCT-1995; 95AU-0006275.  
 PR (UYME ) UNIV MELBOURNE  
 XX (VICT-) VICTORIAN DAIRY IND AUTHORITY.  
 PA Bhogal PS, Reynolds EC, Slakeski N;  
 PI WPI: 1997-272112/24.  
 DR N-PSDB; T78851.  
 XX PT New antigenic protein complex from Porphyromonas gingivalis -  
 PT comprising Arg- and Lys- specific thiol endo-peptidase(s), used in  
 PT the detection, prevention and treatment of periodontal disease  
 XX PS Example 1; Fig 9b; 68pp; English.  
 CC A PrtR-PrtK cell surface protein of Porphyromonas gingivalis (PG)  
 CC comprises a 300 kDa complex composed of a 48 kDa lysine-specific  
 CC thiol protease and 39, 15 and 44 kDa adhesins encoded by the prt  
 CC gene (T78851), and a 45 kDa arginine-specific thiol protease and  
 CC 44, 15, 17 and 27 kDa adhesins (see W24786) encoded by the prt  
 CC gene (T78850). A claimed antigenic complex comprises at least one  
 CC multimeric protein complex of PrtR and PrtK each containing at  
 CC least one adhesin domain, the complex having a mol.wt. of over 200  
 CC kDa, and preferably comprises all 9 proteins of the PrtR-PrtK  
 CC complex (see also W24780-85). It can be used in a claimed  
 CC composition to elicit an immune response directed against PG, and  
 CC in a claimed method of reducing the prospect of PG infection and/or  
 CC severity of disease. Antibodies directed against the complex are  
 CC claimed for use in treating PG infection. Unlike whole PG cells or  
 CC other previously prepared antigens based on fibrillar or the  
 CC capsule, the PrtR-PrtK complex or component parts are safe and  
 CC effective antigens.  
 XX SQ Sequence 1732 AA;  
 OY 2 SFPAIGDTTAQVPEFI 17  
 Db 1658 ntsladvtaqkpytl 1673  
 RESULT 10  
 W69487  
 ID W69487 standard; Protein; 1732 AA.  
 AC W69487;  
 XX 22-DEC-1998 (first entry)  
 DE Haemagglutinin protein prtP.  
 XX Haemagglutinin protein; periodontal disease; vaccine; prtP.  
 KW Porphyromonas gingivalis.  
 XX OS US5824791-A.  
 XX PN 20-OCT-1998.  
 XX PD 11-DEC-1995; 95US-0570311.  
 XX PF 11-DEC-1995; 95US-0570311.  
 XX PR 11-DEC-1995; 95US-0570311.  
 XX PR 08-SEP-1988; 88US-0241640.  
 XX PR 25-JAN-1991; 91US-0647119.  
 XX PR 09-DEC-1994; 94US-0353485.  
 XX PA (UABR-) UAB RES FOUND.

PA (UYFL ) UNIV FLORIDA.  
 XX Han N, Lantz M, Lepine G, Patti JM, Progulskie-Fox A;  
 PI Tumwasorn S;  
 XX WPI: 1998-582627/49.  
 DR N-PSDB: V58874.  
 XX Isolated Porphyromonas gingivalis genes - encoding haemagglutinin  
 PT and/or protease poly:peptide(s))  
 XX Claim 1; Column 69-84; 101pp; English.  
 PS  
 CC This sequence is encoded by a Porphyromonas gingivalis gene of the  
 CC invention. This sequence represents the prp haemagglutinin protein. The  
 CC polypeptides are used to produce antibodies to organisms associated with  
 CC periodontal disease. The antibodies are also used in purification and  
 CC identification procedures. The genes and polypeptides are used as  
 CC vaccines against periodontal disease.  
 CC  
 SQ Sequence 1732 AA;

Query Match 46.0%; Score 40; DB 19; Length 1732;  
 Best Local Similarity 37.5%; Pred. No. 2e+02; Mismatches 4; Indels 0; Gaps 0;  
 Matches 6; Conservative 6;

OY 2 SFPAIGDTTAQVPEFI 17  
 : : : | | | | : : :  
 Db 1658 nitsladvtaqkpytl 1673

## RESULT 11

R96030 ID R96030 standard; Protein; 2628 AA.

XX AC R96030;

DT 04-SEP-1996 (first entry)

XX DE P. gingivalis haemagglutinin hgaA.

XX KW Haemagglutinin; hgaA; periodontal disease; vaccine; antibody.

OS Porphyromonas gingivalis strain 381.

XX XX Key Location/Qualifiers

FT Peptide 5..21

XX W09617936-A2.

XX PD 13-JUN-1996.

XX PF 11-DEC-1995; 95WO-US16108.

XX PR 09-DEC-1994; 94US-0353485.

XX PA (UABR-) UAB RES FOUND.

XX PA (UYFL ) UNIV FLORIDA.

XX PI Han N, Lantz M, Lepine G, Patti JM, Progulskie-Fox A;

XX PI Tumwasorn S;

XX WPI: 1996-287181/29.

XX DR N-PSDB: T30654.

XX Porphyromonas gingivalis genes and proteins - used in the detection

XX and vaccination against periodontal disease

XX Claim 6; Page 93-101; 153pp; English.

XX P. gingivalis 381 haemagglutinin hgaA (R96030) was identified as

CC the product of the hgaA gene (T30654) isolated as an EcoRV  
 CC fragment of genomic DNA. The haemagglutinin, or portions of it  
 CC (see also R96021-24), can be obtained from transformed host cells and  
 CC used as a vaccine to protect humans or animals against periodontal  
 CC disease. Expression in *Salmonella* cells allows production of a live  
 CC vaccine. The haemagglutinin can also be used to detect the  
 CC presence of anti-P. gingivalis antibodies and to raise monoclonal  
 CC antibodies for diagnostic application.  
 CC  
 SQ Sequence 2628 AA;

Query Match 46.0%; Score 40; DB 17; Length 2628;  
 Best Local Similarity 37.5%; Pred. No. 3.3e+02;  
 Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 2 SFPAIGDTTAQVPEFI 17  
 : : : | | | | : : :  
 Db 2554 nitsladvtaqkpytl 2569

## RESULT 12

W69488 ID W69488 standard; Protein; 2628 AA.

XX AC W69488;

DT 22-DEC-1998 (first entry)

XX DE Haemagglutinin protein hgaA.

XX KW Haemagglutinin protein; periodontal disease; vaccine; hgaA.

OS Porphyromonas gingivalis.

XX XX US5824791-A.

XX PD 20-OCT-1998.

XX PE 11-DEC-1995; 95US-0570311.

XX PR 11-DEC-1995; 95US-0570311.

XX PR 08-SEP-1988; 88US-0241640.

XX PR 25-JAN-1991; 91US-0647119.

XX PR 09-DEC-1994; 94US-0353485.

XX PA (UABR-) UAB RES FOUND.

XX PA (UYFL ) UNIV FLORIDA.

XX PI Han N, Lantz M, Lepine G, Patti JM, Progulskie-Fox A;

XX PI Tumwasorn S;

XX WPI: 1998-582627/49.

XX DR N-PSDB: V58875.

XX Isolated Porphyromonas gingivalis genes - encoding haemagglutinin

XX and/or protease poly:peptide(s))

XX Claim 1; Column 91-110; 101pp; English.

XX This sequence is encoded by a Porphyromonas gingivalis gene of the  
 CC invention. This sequence represents the hgaA haemagglutinin protein. The  
 CC polypeptides are used to produce antibodies to organisms associated with  
 CC periodontal disease. The antibodies are also used in purification and  
 CC identification procedures. The genes and polypeptides are used as  
 CC vaccines against periodontal disease.  
 CC  
 SQ Sequence 2628 AA;

Query Match 46.0%; Score 40; DB 19; Length 2628;  
 Best Local Similarity 37.5%; Pred. No. 3.3e+02;  
 Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 2 SFTAGDTTAYVPSIV 17  
 ID B41673 standard; Protein; 248 AA.  
 AC B41673;  
 XX 08-FEB-2001 (first entry)  
 DE Human ORFX ORF1437 polypeptide sequence SEQ ID NO:2874.

Human: open reading frame: ORFX; detection: cytostatic; hepatotropic;  
 vulnerability: antiparkinsonian; nontropic; neuroprotective;  
 immunosuppressant; osteopathic; antiallergic; immunosuppressant; cardiant;  
 hypotensive; dermatological; immunosuppressive; antidiabetic;  
 antiviral; antibacterial; antifungal; antirheumatic; antichyroid;  
 antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 neurodegenerative disease; osteoarthritis; graft vs host disease;  
 cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 cholesterol ester storage; systemic lupus erythematosus; infection;  
 severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 thrombosis; contraceptive.

Homo sapiens.  
 MO200058473-A2.  
 PD 05-OCT-2000.  
 PF 31-MAR-2000; 2000MO-US08621.  
 PR 31-MAR-1999; 9905-0127607.  
 PR 02-APR-1999; 9905-0127636.  
 PR 05-APR-1999; 9905-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 PA (CURA-) CURAGEN CORP.  
 PI Shinkets RA, Leach M;  
 DR WPI; 2000-602362/57.  
 DR N-PSDB; C75882.

Novel nucleic acids and peptides derived from open reading frame X,  
 useful for treating e.g. cancers, proliferative disorders,  
 neurodegenerative disorders and cardiovascular disease -  
 Claim 11; Page 2112; 5507pp: English.

C7446 to C7606 encode the proteins given in B40337 to B43397, which  
 represent the human ORFX open reading frames 1 to 3161. The ORFX  
 sequences have activities such as: cytostatic; hepatotropic; vulnerability;  
 antiparkinsonian; nontropic; neuroprotective; osteopathic;  
 anticonvulsant; antiallergic; immunosuppressant; immunostimulant;  
 cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
 dermatological; immunosuppressive; antirheumatic; antichyroid;  
 antiviral; antifungal; antirheumatic; antidiabetic; antianemic;  
 sequences can be used for determining the presence of or predisposition  
 to, or preventing or treating pathological conditions associated with an  
 ORFX-associated disorder. The nucleic acids can be used to express ORFX  
 proteins in gene therapy vectors. The proteins and nucleic acids may be  
 used to treat cancers, proliferative disorders, neurodegenerative  
 disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
 diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
 storage, systemic lupus erythematosus, severe combined immunodeficiency

CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
 disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
 cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to  
 enhance coagulation; to inhibit thrombosis; and as a contraceptive.  
 CC  
 XX

SO Sequence 248 AA;  
 Query Match 44.8%; Score 39; DB 21; Length 248;  
 Best Local Similarity 50.0%; Pred. No. 30;  
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
 OY 5 AIGDTTAYVPSIV 18  
 DB 183 alastlimvfgll 196

RESULT 14  
 ID B34831 standard; Protein; 332 AA.  
 AC B34831;  
 XX 26-JAN-2001 (first entry)  
 DE Gene 27 human secreted protein homologous amino acid sequence #119.

Human; secreted protein; diagnosis; cytostatic; immunosuppressive;  
 nontropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
 antidiabetic; antiinflammatory; antilucer; vulnerability; anticonvulsant;  
 antibacterial; antifungal; antiparkinsonian; gene therapy;  
 cancer; immune disorder; cardiovascular disorder; wound healing;  
 neurological disease; infectious disease; chromosome identification.  
 XX  
 XX  
 OS Synchocystis sp.  
 PN MO200058356-A1.  
 PD 05-OCT-2000.  
 PF 22-MAR-2000; 2000MO-US07535.  
 PR 26-MAR-1999; 9905-0126511.  
 PR 17-DEC-1999; 9905-0172413.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Ruben SM, Komatsoulis G;  
 DR WPI; 2000-594639/56.

Fifty nucleic acid molecules encoding human secreted proteins, useful  
 in the prevention, treatment and diagnosis of cancer, immune disorders,  
 cardiovascular disorders and neurological diseases -  
 Disclosure; Page 404-405; 425pp: English.

The polynucleotide sequences given in C59966 to C60015 encode the human  
 secreted proteins given in B34773 to B34822. B34823 to B34852 represent  
 human secreted polypeptide sequences and proteins homologous to them,  
 which are given in the exemplification of the present invention. Human  
 secreted proteins have activities based on the tissues and cells the  
 genes are expressed in. Examples of activities include: cytostatic;  
 immunosuppressive; nontropic; neuroprotective; antiviral; antiallergic;  
 hepatotropic; antidiabetic; antiparkinsonian; antilucer; vulnerability;  
 anticonvulsant; antibacterial; antifungal; antiparkinsonian; cardiant;  
 The polynucleotides and polypeptides are useful for preventing, treating  
 or ameliorating a medical condition in e.g. humans, mice, rabbits,  
 goats, horses, cats, dogs, chickens or sheep. The polypeptides can also  
 be used as a food additive or preservative to increase or decrease  
 storage capabilities. The polynucleotides are useful for chromosome  
 identification. They are also useful as probes for diagnosing a disorder  
 related to the female reproductive system, particularly breast and/or

sequence	332	AA
----------	-----	----

QY 1 ASFTAI GDTTA QVPFSIV 18  
101111 .11.1:

Db 173 atftvngktevtlpydi 190

RESULT 15

ID B27624 standard; Protein; 393 AA.

AC B27624-

DT 02-FEB-2001 (first entry)

DE Human secreted protein BLAST search SEQ ID NO: 125

AA Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KM Cytostatic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
 KM antiallergic; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KM vulnery; anticancer; cancer; immune disorder; cardiovascular disorder  
 KM cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder  
 KM neurological disease; infection; human; secreted protein.

OS Homo sapiens.

PN WO2000055175-A1.

PD 21-SEP-2000.

PF 09-MAR-2000; 2000WO-US06049.

PR 12-MAR-1999; 99US-0124144.

PR 03-DEC-1999; 99US-0168667.

PA (HUMA-) HUMAN GENOME SCI INC

PI Rosen CA, Ruben SM, Komatsoulis G;

DR WPI; 2000-638175/61.

Isolated nucleic acid molecule encoding a human secreted protein is used in preventing, treating or ameliorating a medical condition -

PS Disclosure; Page 410-412; 428pp; English..

The invention relates to the isolation of genes C589980-659039 encoding the human secreted proteins B27560-B27609. This sequence represents a peptide fragment homologous to the protein encoded by the gene given in the descriptor line. The sequence is a search result from a BLASTX homology search. The genes and proteins are useful for preventing, ameliorating or treating medical conditions, e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer, and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver CC lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid

CC arthritis and ulcerative colitis; (c) cardiovascular disorders such  
CC as myocardial ischemias; (d) wound healing; (e) neurological diseases  
CC c.g. cerebral anoxia and epilepsy; and (f) infectious diseases such  
CC as viral, bacterial, fungal and parasitic infections.

Sequence	393 AA
5Q	

Query Match	44.8%;	Score 39;	DB 21;	Length 393;
Best Local Similarity	38.9%;	Pred. No. 52;		
Matches	7; Conservative	4; Mismatches	7; Indels	0; Gaps

```
QY      1 ASFTAIGDTTAQVPESIV 18
         ||| | :| :|
Db      234 atftvngktevltlpydi 251
```

Db 234 atftvngktevtlpydli 251

Search completed: June 20, 2001, 13:32:02  
Job time: 40 sec



